

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE CIENCIAS MATEMÁTICAS**

**Departamento de Estadística e Investigación Operativa I**



**TESIS DOCTORAL**

**Modelos marginales: nuevos procedimientos de  
inferencia para datos longitudinales**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Rosa Alonso Sanz**

Directora:

**María del Carmen Pardo Llorente**

**Madrid, 2011**

**ISBN: 978-84-694-6643-8**

© Rosa Alonso Sanz, 2011

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

Departamento de Estadística e Investigación Operativa I



**Modelos Marginales:  
Nuevos Procedimientos de Inferencia  
para Datos Longitudinales**

Tesis doctoral realizada por:

**D<sup>a</sup>. Rosa Alonso Sanz**

Bajo la dirección de:

**Dra. D<sup>a</sup>. María del Carmen Pardo Llorente**

FEBRERO 2011



D<sup>a</sup>. María del Carmen Pardo Llorente, Profesora Titular del Departamento de Estadística e Investigación Operativa de la Universidad Complutense de Madrid, **AUTORIZA:**

La presentación de la Tesis Doctoral titulada ***Modelos Marginales: Nuevos Procedimientos de Inferencia para Datos Longitudinales***, realizada por D<sup>a</sup> Rosa Alonso Sanz bajo mi dirección en el Departamento de Estadística e Investigación Operativa y que presenta para la obtención del grado de Doctor, por la Universidad Complutense de Madrid.

Madrid a 18 de Febrero de 2011.

María del Carmen Pardo Llorente

*A mi Madre y a mis Hermanos*

*A la memoria de mi Padre, de mi amigo Carlos y de mi prima Tena*

# Agradecimientos

Enumerar en una sola página mi agradecimiento a todas las personas que me han apoyado en la realización de esta tesis es muy complicado. En cualquier caso, agradezco a todos los que de una forma u otra me han ayudado a conseguirlo y si me dejo a alguien que sepa que es por despiste.

Debo expresar mi agradecimiento al Ministerio de Ciencia e Innovación por haber financiado esta tesis mediante la beca FPI de referencia **BES-2007-14901** y al profesor Leandro Pardo, investigador principal de este proyecto. De una forma especial quiero agradecer a la profesora María del Carmen Pardo directora de mi tesis, el haberme dado la oportunidad de trabajar con ella y haber hecho posible la realización de esta tesis doctoral. Le agradezco también su paciencia, su ilusión, su energía y todo su esfuerzo realizado para hacer posible esta tesis doctoral. Muchas gracias Carmen.

Mi agradecimiento a todos los compañeros del departamento de Estadística e Investigación Operativa I de la Universidad Complutense de Madrid, en especial a la profesora Begoña Vitoriano por su amistad, sus consejos y por dejarme participar en su proyecto en Mozambique, a Alicia y a María secretarías del departamento por apoyarme y comprenderme. Un agradecimiento muy especial se merecen mis compañeros de despacho Grego, Tingo y Martín y mis compañeros de doctorado en especial Javi y Paola gracias por animarme y ayudarme siempre que os lo he pedido y Samuel y Karina.

Mi agradecimiento a todos mis compañeros de la Facultad de Biología de la Universidad Francisco de Vitoria por vuestros ánimos y amistad.

Mi más especial agradecimiento está dedicado a mi madre, Tita, por su apoyo, su cariño y su esfuerzo que me han permitido realizar todos mis estudios. A mis hermanos Mariano y Blanca, a mi cuñada Sonia y a mis sobrinas Marta y Cristina por su apoyo, sus ánimos y su cariño. Gracias también a mis Tíos Mari, Gloria, Lucas y Matías y a todos mis primos, en especial a Raquel, por quererme y animarme.

Mi agradecimiento a todos mis amigos, las Apaños y en especial a la Pony, a los Chinchés, a la familia Colorín en especial a mi ahijado Julian y a Cote gracias por vuestro apoyo, amistad y por soportarme. Y finalmente a Gabi en parte su constante apoyo, cuidados, ayuda, aliento, cariño, ánimo y comprensión me ha permitido cierta tranquilidad en estos últimos momentos de fin de tesis. Gracias Gabi.

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Datos Longitudinales . . . . .	1
1.2. Modelos estadísticos para datos longitudinales . . . . .	3
1.2.1. Modelos Condicionales . . . . .	4
1.2.2. Modelos Marginales . . . . .	6
1.2.3. Modelos con parámetros específicos para cada individuo . . . . .	7
1.3. Organización de los otros capítulos . . . . .	8
<b>2. Ajuste de modelos marginales por mínima <math>\phi</math>-divergencia</b>	<b>11</b>
2.1. Modelo de Bahadur . . . . .	12
2.2. Formulación del modelo de Dale . . . . .	14
2.2.1. Parametrización Lineal para Modelos CRG . . . . .	16
2.2.2. Ajuste Modelo de Dale por máxima verosimilitud . . . . .	18
2.3. Ajuste del Modelo de Dale por mínima $\phi$ -divergencia . . . . .	20
2.3.1. Comportamiento asintótico del estimador de mínima $\phi$ -divergencia	24
2.4. Aplicación numérica . . . . .	32
2.5. Estudio de simulación . . . . .	53
<b>3. Extensión de las Ecuaciones de Estimación Generalizadas</b>	<b>75</b>
3.1. Modelo Lineal Generalizado . . . . .	76
3.2. Metodología de las Ecuaciones de Estimación Generalizadas . . . . .	78
3.2.1. Elección de la matriz de correlación de trabajo . . . . .	81
3.2.2. Algoritmo para encontrar la solución EEG . . . . .	85

3.3. Procedimientos alternativos a las EEG . . . . .	87
3.3.1. Nuevos procedimientos basados en medidas de divergencia . . . . .	88
3.4. Ejemplos . . . . .	92
3.4.1. Ejemplo 1. Tabla de Contingencia . . . . .	93
3.4.2. Ejemplo 2. Insuficiencia respiratoria . . . . .	98
3.4.3. Ejemplo 3: Uso de las EEG para Datos de Conteo . . . . .	103
3.5. Estudio de simulación . . . . .	107
3.5.1. Modelo para variables binarias correlacionadas . . . . .	108
3.5.2. Modelo para variables Poisson correlacionadas . . . . .	121
<b>4. Medidas de Diagnóstico</b>	<b>131</b>
4.1. Introducción . . . . .	131
4.2. Gráficos cuantil-cuantil basados en residuos $\phi$ -divergencia . . . . .	133
4.3. Estudio de Simulación . . . . .	134
4.3.1. Para variables binarias . . . . .	135
4.3.2. Para variables Poisson . . . . .	141
4.4. Medidas basadas en el volumen de elipsoides de confianza . . . . .	147
4.5. Ejemplos . . . . .	151
4.5.1. Ejemplo respuesta Poisson . . . . .	151
4.5.2. Ejemplo respuesta binaria . . . . .	166
<b>5. Conclusiones y líneas futuras de investigación</b>	<b>173</b>
<b>Anexo</b>	<b>179</b>
<b>Bibliografía</b>	<b>197</b>

# Capítulo 1

## Introducción

### 1.1. Datos Longitudinales

La Estadística Aplicada tiene entre sus objetivos principales construir modelos que expliquen las relaciones entre los datos recopilados. En este manuscrito se está interesado en situaciones en las que el objetivo es explicar como una variable de particular interés (variable respuesta) está relacionada con un conjunto de variables (variables explicativas). En el contexto clásico, una de las hipótesis fundamentales, a la hora de la modelización, era la independencia entre observaciones. Sin embargo, muchos estudios tienen diseños que implican datos correlacionados. Se considera este término en un sentido genérico, que engloba estructuras como observaciones multivariantes, datos cluster, medidas repetidas, datos longitudinales y datos correlacionados.

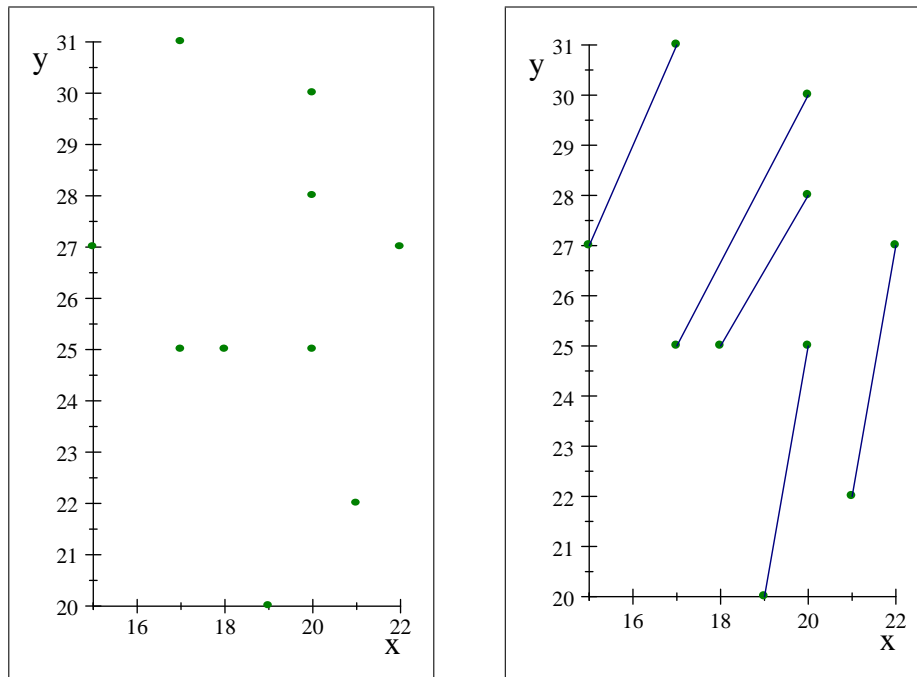
Cuando la misma característica es medida repetidamente en el tiempo y el tiempo a su vez es, al menos en parte, objeto de estudio, se tienen datos longitudinales. En este sentido, son diseños mixtos, caracterizados por considerar simultáneamente dos o más dimensiones de análisis, en los que el tiempo constituye una de ellas. El análisis longitudinal presenta importantes ventajas respecto a otros diseños. La más importante es que sus diseños son más eficientes, al permitir distinguir variaciones inter e intraindividuales.

En muchas ocasiones los científicos pueden responder a las mismas cuestiones mediante un estudio longitudinal o mediante un estudio transeccional, aunque puede ocurrir



lo que se muestra a continuación con un ejemplo acerca de un estudio entre la edad y la comprensión lectora.

Figura 1.1. Estudio Transeccional (izquierda), Estudio Longitudinal (derecha).



x= edad, y=comprensión lectora.

El gráfico de la izquierda de la Figura 1.1 muestra la edad frente a la comprensión lectora en un hipotético estudio transeccional. En ella parece que la comprensión lectora es menor en chicos más mayores. En el gráfico de la derecha de la Figura 1.1 suponemos el mismo conjunto de datos pero obtenido en un estudio longitudinal en el que cada individuo ha sido medido dos veces. Observamos como la habilidad lectora mejora con el tiempo para cada individuo, aunque los niños más pequeños empezaron con un nivel más alto de lectura.

Los estudios longitudinales son estudios en los que la respuesta de cada individuo se observa en dos o más ocasiones y el tiempo en si mismo, al menos en parte, es sujeto de investigación. Estos estudios representan una de las principales estrategias de investigación

empleadas en las Ciencias Sociales y Medicina, Goldstein [46]; Nesselroade y Baltes [74]. También tienen aplicaciones a la Economía y a la Bioestadística. Agresti [2], Diggle et al. [39]. Baltes y Nesselroade [13] investigan la historia de los datos longitudinales y sus métodos asociados. Describen la investigación longitudinal como una variedad de métodos conectados por la idea de que la entidad bajo investigación se observa repetidamente ya que existe y se desarrolla a lo largo del tiempo. Además, describen investigaciones longitudinales en el siglo XIX. Toon [109] cita la encuesta de los presupuestos de Engel en 1857, examinando como la cantidad de dinero gastada en comida cambia como una función del tiempo, como probablemente el primer ejemplo de un estudio involucrando medidas repetidas del mismo conjunto de individuos.

## 1.2. Modelos estadísticos para datos longitudinales

Los modelos estadísticos a utilizar dependen del tipo de variable-respuesta analizada. Los métodos para variable respuesta continua bajo suposiciones de normalidad son los que han tenido un mayor desarrollo. En concreto, el modelo lineal mixto (Laird y Ware [59], Verbeke y Moleberghs [111]) ha jugado un importante papel en extender el modelo lineal general para tratar con datos continuos correlacionados. Debido a las propiedades de la distribución normal multivariante, su teoría e implementación se simplifica bastante. Programas de software tales como el procedimiento MIXED de SAS [63], son ampliamente utilizados para ajustar esta clase de modelos y han facilitado la difusión de esta metodología. Sin embargo, cuando la variable respuesta es discreta o categórica se dispone de menos técnicas debido a la falta de un análogo discreto a la distribución normal multivariante.

En los últimos 30 años, el desarrollo de métodos estadísticos para datos longitudinales ha sido considerable. Desde luego siempre hay un desfase entre los últimos avances publicados en las revistas y su aplicación a problemas reales. La razón de ésto, al menos en parte, es que su implementación en el software estándar es más lento. Sin embargo, la introducción reciente de nuevos programas para analizar datos longitudinales ha hecho que estos métodos sean más accesibles a investigadores. Por otra parte, puesto que el software estadístico está en continuo desarrollo, los métodos más modernos para análisis

longitudinales serán utilizados más ampliamente y en un gran número de disciplinas.

De acuerdo con Fahrmeir y Tutz [40], Diggle et al. [39] y Molenberghs y Verbeke [69] se pueden distinguir tres amplias familias de modelos para datos longitudinales:

1. modelos especificados condicionalmente,
2. modelos marginales,
3. modelos específicos del individuo.

La elección práctica de la familia de modelos depende principalmente de la(s) cuestión(es) a ser contestadas. En modelos especificados condicionalmente la esperanza de la respuesta en un tiempo dado se modeliza en términos de las respuestas en otros tiempos así como de las variables explicativas. En modelos marginales, las variables explicativas están directamente relacionadas con las esperanzas marginales. Los modelos específicos de los individuos se diferencian de los anteriores en la inclusión de parámetros que son específicos de los individuos o unidades independientes. La selección del método para ajustar el modelo, no debería depender únicamente de las suposiciones que el investigador quiera hacer, sino también de la disponibilidad de algoritmos computacionales. Si se pueden especificar por completo las probabilidades conjuntas, se adoptarán métodos de máxima verosimilitud. Sin embargo, si únicamente se dispone de una descripción parcial en términos de probabilidades marginales o condicionadas se deben utilizar métodos tales como las ecuaciones de estimación generalizadas o métodos de pseudoverosimilitud.

### 1.2.1. Modelos Condicionales

En un modelo condicional los parámetros describen características (esperanza, probabilidad, odds ratios, logit,...) de un conjunto de respuestas, dados los valores de las variables explicativas (Cox [29]). El ejemplo más conocido es el modelo loglineal. Debido a la popularidad de los modelos marginales, especialmente de las ecuaciones de estimación generalizadas y de los modelos de efectos-aleatorios para medidas repetidas, a los modelos condicionales se les ha prestado relativamente poca atención dentro de este contexto. Diggle et al. [38] criticaron esta aproximación debido a que la interpretación de un parámetro de efecto (por ejemplo, la evolución del tiempo o el efecto del tratamiento) de

una respuesta se condiciona a otras respuestas para el mismo individuo, a respuestas de otros individuos, y al número de respuestas repetidas. Molenberghs y Ryan [71] discuten las ventajas de los modelos condicionales y como se pueden solucionar las desventajas. Construyeron la distribución conjunta para repuestas cluster binarias multivariantes, basada en un modelo de familia exponencial multivariante. Otra aproximación, también basada en la familia exponencial, se puede ver en Fitzmaurice, Laird y Tosteson [43]. Una ventaja de estas aproximaciones es que, bajo una especificación correcta del modelo, pueden ganar eficiencia frente a otros modelos tales como los métodos de las ecuaciones de estimación generalizadas.

A pesar de las críticas, los modelos condicionales gozan de una cierta popularidad en algunas áreas tales como el análisis multivariante, en particular aplicado a las ciencias sociales. Quizás la razón más importante para esta popularidad es la conveniencia matemática de los modelos loglineales. Sin embargo, la interpretación de los parámetros es difícil. También, la estimación de máxima verosimilitud puede resultar prohibitiva debido a la carga computacional.

Un caso particular son los modelos de transición o Markov. En modelos de transición, la distribución condicionada de cada respuesta se expresa como una función explícita de las respuestas pasadas y las variables explicativas. La dependencia entre las medidas repetidas se debe a la influencia de los valores pasados de la respuesta en la observación actual. Cuando la variable respuesta es discreta, estos modelos se conocen como modelos de cadenas de Markov. Existe una extensa bibliografía del uso de cadenas de Markov para modelizar datos longitudinales discretos igualmente espaciados con un número finito de estados o categorías (por ejemplo, Anderson y Goodman [7] ; Cox [28]; Billingsley [19]). En el modelo más simple para datos longitudinales, una cadena de Markov de primer orden, las probabilidades de transición se suponen iguales para cada intervalo de tiempo. Puesto que, en general, los modelos de transición han sido desarrollados para medidas repetidas que están separadas por igual en el tiempo, estos modelos son más difíciles de aplicar cuando hay datos “missing”. Además, la interpretación de los parámetros cambia.

### 1.2.2. Modelos Marginales

En un modelo marginal o modelo de población -promedio, los parámetros caracterizan la esperanza marginal (por ejemplo, la probabilidad marginal de un suceso en un punto en el tiempo es dado cuando la respuesta es binaria) de un subconjunto de las respuestas, sin condicionar otras respuestas.

De hecho el desarrollo de métodos para analizar respuestas longitudinales continuas se ha llevado a cabo a lo largo de más de un siglo, desde el primer trabajo de modelos con efectos aleatorios del astrónomo británico Airy [4], hasta el trabajo clave sobre modelos lineales con efectos mixtos para datos longitudinales de Laird y Ware [59]. Sin embargo, la mayoría de los avances de métodos para datos longitudinales discretos se concentran en los últimos 25 a 30 años coincidiendo con la disponibilidad a nivel de usuario de ordenadores de alta velocidad.

Cuando la respuesta longitudinal es discreta, los modelos lineales ya no son apropiados para relacionar cambios entre la respuesta media y las variables explicativas. En su lugar, se desarrollaron extensiones de los modelos lineales generalizados (Nelder y Wedderburn, [73]) para datos longitudinales de diferentes formas.

Hay una literatura amplia de estos modelos marginales para respuestas binarias correlacionadas. Por ejemplo, Bahadur [12], Zhao y Prentice [118], proponen estimación de máxima verosimilitud donde se utilizan las correlaciones marginales para tener en cuenta la asociación entre las respuestas. Alternativamente, la asociación también se puede parametrizar en términos de los “odds ratio” marginales, como se puede ver por ejemplo en Dale [36], Liang et al. [62], Lang y Agresti [3], Molenberghs y Lesaffre [70], Glonek y McCullagh [45].

Hay numerosas propuestas de modelos flexibles en la literatura, sin embargo, la máxima verosimilitud puede no ser atractiva debido al excesivo gasto computacional, especialmente cuando tenemos vectores de gran dimensión de datos correlacionados. Como consecuencia de esto, Liang y Zeger [61] han propuesto métodos alternativos como las *ecuaciones de estimación generalizadas* (EEG), en la que sólo se necesita una correcta especificación de la distribución marginal univariante siempre que estemos dispuestos a asumir la estructura de la matriz de “trabajo”. Extensiones de las EEG que permiten

modelizar las asociaciones por pares fueron dadas, por ejemplo, por Prentice [92], Lipsitz et al. [64], Liang et al. [62] y Carey et al. [23] utilizando como medida de asociación bien la correlación bien los “odds ratio”.

Una alternativa a esta aproximación es dada por Cessie y van Houweling [24] que proponen aproximar la verdadera verosimilitud por medio de la función de *pseudo-verosimilitud* (PL) que es más fácil de evaluar y de maximizar.

### 1.2.3. Modelos con parámetros específicos para cada individuo

Los modelos con parámetros específicos para cada individuo se diferencian de los modelos marginales o modelos de población promedio por la inclusión de parámetros específicos de los individuos. Los parámetros de estos modelos para datos no gaussianos describen diferentes tipos de efectos de las variables explicativas en las esperanzas, Neuhaus et al. [75]. Por ejemplo, cuando las respuestas son binarias, los efectos de las variables explicativas en las probabilidades condicionadas están condicionados al nivel del efecto específico del individuo. Esto significa que una unidad de cambio en las variables explicativas se traduce en un cambio apropiado en la probabilidad, manteniendo el nivel del efecto específico del individuo fijo. Los modelos son útiles si estamos interesados en cambios dentro de cada sujeto (Neuhaus et al. [75]). En modelos marginales, tal diferencia en las variables explicativas se traduce en diferencia en la probabilidad respuesta marginal sobre los efectos específicos del individuo. No existen únicamente diferencias en la interpretación entre ambas familias sino también podemos observar diferencias sustanciales cuando comparamos los parámetros estimados por cada una de ellas. La razón clave es que estimamos diferentes parámetros “verdaderos”. Zeger et al. [117] comparan estas 2 aproximaciones en la modelización de datos longitudinales utilizando EEG para estimar los parámetros.

Hay tres formas de tratar estos modelos: (a) como efectos fijos, (b) como efectos aleatorios y (c) condicionando por ellos. La primera parece la más simple pero en muchos casos no es así debido a que el número de parámetros crece proporcionalmente al tamaño muestral, por lo que no son válidos la mayoría de los resultados inferenciales clásicos. La segunda es muy popular y hay varias vías para introducir la aleatoriedad en los parámetros. Algunos autores como Stiratelli et al. [104] y Breslow y Clayton [20] suponen

que el vector de parámetros es normal. Alternativamente, Skellam [103] introduce el modelo beta-binomial. La tercera vía es bien conocida en epidemiología. En particular, la regresión logística condicionada se considera con frecuencia (Breslow y Day [21]).

El debate acerca de las diferentes ventajas de cada una de las familias de modelos continúa. La elección del modelo no dependerá únicamente de la aplicación de interés sino también de los objetivos específicos del análisis. Algunos investigadores han propuesto combinar diferentes aproximaciones para solucionar los problemas de una única familia de modelos. Una primera posibilidad es modificar la parametrización natural de los modelos condicionales para permitir la estimación de máxima verosimilitud de los parámetros de la media marginal (Fitzmaurice y Laird [41], Azzalini [11], Heagerty [50] y Heagerty y Zeger [51]).

### 1.3. Organización de los otros capítulos

El objetivo que ha guiado esta tesis ha sido el de analizar la importancia de las medidas  $\phi$ -divergencia en el estudio de modelos marginales para datos longitudinales.

El Capítulo 2 está dedicado a modelos marginales basados en verosimilitud y se estudiará en detalle el modelo propuesto por Dale [36]. El estimador clásico de máxima verosimilitud, para los parámetros de este modelo, se puede obtener como el valor del espacio paramétrico que minimiza la divergencia de Kullback entre el vector de probabilidades que caracteriza el modelo y el estimador no paramétrico del mismo. En consecuencia al ser la divergencia de Kullback un caso particular de medida de  $\phi$ -divergencia, el estimador de mínima  $\phi$ -divergencia surge de forma natural al sustituir la divergencia de Kullback por la medida de  $\phi$ -divergencia. Se establece que las propiedades asintóticas de los estimadores de mínima  $\phi$ -divergencia son independientes de la función  $\phi$  considerada. En consecuencia, todos tienen el mismo comportamiento asintótico que a su vez coincide con el del estimador de máxima verosimilitud. Se presenta un ejemplo que ilustra la nueva familia de estimadores. Se finaliza este Capítulo 2 con un estudio de simulación para analizar el comportamiento de los estimadores de mínima  $\phi$ -divergencia para los modelos propuestos por Dale.

En algunos casos cuando se tienen vectores de datos correlacionados de grandes di-

mentaciones la máxima verosimilitud puede ser prohibitiva debido a los requerimientos computacionales. Como consecuencia aparecen métodos alternativos como las Ecuaciones de Estimación Generalizadas (EEG) propuestas por Liang y Zeger [61] que suponen que la distribución marginal de la variable respuesta sigue un modelo lineal general (McCullagh y Nelder [68]) y que la asociación entre las observaciones de cada individuo tiene una estructura arbitraria, que denominan correlación de trabajo. En el Capítulo 3, el objetivo es la extensión de las EEG.

El algoritmo para resolver las EEG utiliza el método de los momentos para estimar los parámetros de correlación basándose en los residuos de Pearson. Park et al. [84] propusieron utilizar otros residuos para llevar a cabo dicha estimación. En concreto, propone los residuos desviación y de Anscombe. Pero los residuos desviación no son más que la divergencia de Kullback entre lo observado y lo esperado y los residuos de Pearson no son más que la divergencia de Kagan entre lo observado y lo esperado. Por tanto, una alternativa sería utilizar en lugar de los residuos basados en estas divergencias cualquier otra divergencia. En concreto, en el Capítulo 3, se propone utilizar residuos basados en la familia de  $\phi$ -divergencias que contienen como caso particular no sólo a las divergencias de Kullback y Kagan sino a otras muchas conocidas. Se presentan ejemplos que abarcan diferentes situaciones. Además, se lleva a cabo un amplio estudio de simulación para analizar el comportamiento de los nuevos procedimientos de estimación para muestras de pequeñas a moderadas.

El Capítulo 4 está dedicado al estudio de técnicas específicas de diagnóstico para detectar observaciones influyentes. En primer lugar, basados en la familia de residuos  $\phi$ -divergencia definida en el capítulo anterior y que generaliza tanto los residuos clásicos de Pearson como los de desviación usualmente utilizados para modelos para datos longitudinales se propone un gráfico cuantil-cuantil para detectar “outlying”. Se realiza un estudio de simulación en el que se comparan diferentes gráficos cuantil-cuantil.

En segundo lugar, puesto que la mayoría de los diagnósticos para EEG son generalizaciones de los diagnósticos para modelos lineales generalizados (Pregibon [90], Williams [114], McCullagh y Nelder [68]). Y de hecho, éstas se reducen a las medidas de influencia conocidas en regresión lineal y que se pueden encontrar en Cook y Weisberg [27], Belsey et al. [18]. En este capítulo, se definen medidas basadas en el cambio del volumen del



elipsoide de confianza. Estas medidas se comparan con las medidas de influencia definidas por Preisser y Qaqish [91] que están implementadas en SAS a través de dos ejemplos.

Finalmente, en el Capítulo 5 se presentan conclusiones y también se señalan algunos tópicos para investigaciones futuras.

## Capítulo 2

# Ajuste de modelos marginales por mínima $\phi$ -divergencia

En este capítulo, nos centramos en modelos de probabilidad especificados completamente, en contraste con aquellos en los que se especifican únicamente unos pocos momentos de orden bajo, tal y como ocurre con las ecuaciones de estimación generalizadas que serán objeto de estudio del siguiente capítulo. Aunque se complican tanto la teoría como los cálculos, hay al menos dos situaciones en las que es preferible éste camino. Primero puede ser necesaria una modelización cuidadosa de la estructura de asociación. En segundo lugar, se puede estar interesado en la probabilidad conjunta de un número de sucesos. Además algunos de éstos modelos son la base para otros métodos no basados en máxima verosimilitud. Por ejemplo, el modelo probabilístico de Bahadur del que se da una breve pincelada en la Sección 2.1 es la base de las EEG introducidas por Liang y Zeger [61]. Sin embargo, el modelo de Bahadur a pesar de su formulación relativamente sencilla tiene unas limitaciones prácticas que hace que se busquen modelos alternativos cuando se desea el análisis de un modelo marginal basado en verosimilitud. Por ello nosotros nos centraremos en el modelo de Dale [36] que se puede ver como la raíz para el modelo no basado en verosimilitud propuesto por Lipsitz et al. [64].

En este capítulo, se presenta un nuevo procedimiento de estimación basado en la medida  $\phi$ -divergencia para el modelo de Dale. Se estudia sus propiedades asintóticas y se

obtiene su distribución asintótica. Se ilustra el nuevo procedimiento de estimación con un ejemplo y finalmente, se realiza un estudio de simulación para analizar el comportamiento de los estimadores de mínima  $\phi$ -divergencia para los modelos propuestos por Dale.

## 2.1. Modelo de Bahadur

Bahadur [12] introdujo este método, con expresión explícita pero con bastantes problemas computacionales debido a la forma de su espacio paramétrico. El modelo se concibió para datos binarios.

Sea  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{it_i})^T$  un vector de respuestas binarias, donde  $Y_{ij}$  es la respuesta del sujeto  $i$  en el tiempo o condición  $j$ , e indica si el evento en estudio ha ocurrido o no.

La distribución marginal de  $Y_{ij}$  es Bernoulli con

$$E(Y_{ij}) = \Pr(Y_{ij} = 1) \equiv \pi_{ij}.$$

Para describir la asociación, se necesita caracterizar la probabilidad conjunta de dos sucesos

$$P(Y_{ij1} = 1, Y_{ij2} = 1) = E(Y_{ij1}Y_{ij2}) \equiv \pi_{ij1j2}.$$

Esta probabilidad de éxito de dos medidas del mismo individuo pueden ser modeladas tanto en función de las probabilidades marginales como en función de un parámetro de asociación, siendo éste el coeficiente de correlación marginal del modelo de Bahadur.

Definimos el coeficiente de correlación marginal de la siguiente manera

$$\rho_{ij1j2} = \text{Corr}(Y_{ij1}, Y_{ij2}) \equiv \frac{\pi_{ij1j2} - \pi_{ij1}\pi_{ij2}}{[\pi_{ij1}(1 - \pi_{ij1})\pi_{ij2}(1 - \pi_{ij2})]^{1/2}} \quad (2.1)$$

En términos de este parámetro de asociación, la probabilidad conjunta  $\pi_{ij1j2}$  puede ser escrita como

$$\pi_{ij1j2} = \pi_{ij1}\pi_{ij2} + \rho_{ij1j2} [\pi_{ij1}(1 - \pi_{ij1})\pi_{ij2}(1 - \pi_{ij2})]^{1/2}. \quad (2.2)$$

Por lo tanto, dado el coeficiente de correlación marginal  $\rho_{ij1j2}$  y las probabilidades  $\pi_{ij1}$  y  $\pi_{ij2}$ , la probabilidad conjunta  $\pi_{ij1j2}$  puede ser calculada con facilidad.

Se tienen que especificar ahora el primer y segundo momento de la distribución. Sin embargo, una aproximación basada en la verosimilitud necesita la representación completa de la probabilidad conjunta del vector de respuestas binarias para cada individuo.

La distribución conjunta de  $f(\mathbf{y})$  de  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{it_i})^T$  es multinomial. Bahadur, utilizó a parte del coeficiente de correlación de orden 2, los coeficientes de correlación de orden 3 y de orden superior para especificar completamente la distribución conjunta. Con este fin, sean

$$\varepsilon_{ij} = \frac{Y_{ij} - \pi_{ij}}{\sqrt{\pi_{ij}(1 - \pi_{ij})}} \quad y \quad e_{ij} = \frac{y_{ij} - \pi_{ij}}{\sqrt{\pi_{ij}(1 - \pi_{ij})}} \quad (2.3)$$

donde  $y_{ij}$  es el valor actual de la variable de respuesta binaria  $Y_{ij}$ .

Sean

$$\begin{aligned} \rho_{ij_1 j_2} &= E(\varepsilon_{ij_1} \varepsilon_{ij_2}), \\ \rho_{ij_1 j_2 j_3} &= E(\varepsilon_{ij_1} \varepsilon_{ij_2} \varepsilon_{ij_3}), \\ &\vdots, \\ \rho_{i12\dots t_i} &= E(\varepsilon_{i1} \varepsilon_{i2} \dots \varepsilon_{it_i}). \end{aligned}$$

Entonces, el modelo general de Bahadur puede ser representado por la expresión

$$f(\mathbf{y}_i) = f_1(\mathbf{y}_i) c(\mathbf{y}_i),$$

donde

$$f_1(\mathbf{y}_i) = \prod_{j=1}^{t_i} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}}$$

y

$$\begin{aligned} c(\mathbf{y}_i) = & 1 + \sum_{j_1 < j_2} \rho_{ij_1 j_2} e_{ij_1} e_{ij_2} + \sum_{j_1 < j_2 < j_3} \rho_{ij_1 j_2 j_3} e_{ij_1} e_{ij_2} e_{ij_3} \\ & + \dots + \rho_{i12\dots t_i} e_{i1} e_{i2} \dots e_{it_i}. \end{aligned}$$

Es decir, la función de densidad es el producto de  $f_1(\mathbf{y}_i)$  y de un factor de corrección  $c(\mathbf{y}_i)$ .

Este desarrollo tiene la propiedad, compartida con el modelo probit multivariante y otros modelos marginales, de ser reproductible en el sentido de que el mismo modelo se mantenga para cualquier subconjunto del vector de respuestas. Además, las probabilidades multinomiales para el vector de respuestas binarias son relativamente fáciles de obtener dados los parámetros del modelo. Kupper y Haseman [58] y Altham [6] propusieron aplicaciones de este modelo aunque con estructuras de correlación dos a dos muy simples y suponiendo que los términos de órdenes altos son cero. Lo que ha hecho que este modelo haya tenido un uso limitado en datos longitudinales ha sido su parametrización de las asociaciones de órdenes altos en términos de los parámetros de correlación. De

hecho, para datos discretos, el modelo de Bahadur requiere un conjunto complicado de restricciones de desigualdad sobre los parámetros del modelo que hacen muy difícil la maximización de la verosimilitud. Excepto para situaciones sencillas con un número pequeño de medidas repetidas, el modelo de Bahadur no se ha utilizado mucho para datos longitudinales.

Debido a las restricciones de las correlaciones, se han propuesto modelos multinomiales alternativos para la distribución conjunta del vector de respuestas discretas donde la asociación dentro de cada sujeto se parametriza en términos de otras métricas de asociación. Por ejemplo, algunos autores ya citados en la Sección 1.2.2 propusieron el método de máxima verosimilitud pero parametrizando los momentos de órdenes altos en términos de “odds ratios” marginales.

## 2.2. Formulación del modelo de Dale

No siempre es posible partir de la función de masa de probabilidad multivariante como en el modelo de Bahadur. En la mayoría de los casos se empieza con las marginales univariantes sobre las que se supone una estructura de asociación de los órdenes segundo y superior hasta completar la especificación del modelo. En el modelo de Dale, que se presenta a continuación, la estructura de asociación considera el “cross ratio” global.

Sea  $\mathbf{Z}(\mathbf{x}) = (Z_1(\mathbf{x}), Z_2(\mathbf{x}))$  un vector de variables aleatorias discretas, dependiendo de un vector  $p$ -dimensional de variables explicativas  $\mathbf{x}$ , cuantitativas o cualitativas. Supongamos que  $Z_1(\mathbf{x})$  toma valores  $1, \dots, r$  de acuerdo con las probabilidades acumuladas

$$\eta_{ix} = \Pr(Z_1(\mathbf{x}) \leq i), i = 1, \dots, r;$$

de la misma manera la función de probabilidad acumulada para  $Z_2(\mathbf{x})$  es

$$\xi_{jx} = \Pr(Z_2(\mathbf{x}) \leq j), j = 1, \dots, c,$$

con

$$\eta_{0x} = \xi_{0x} = 0, \eta_{rx} = \xi_{cx} = 1. \boldsymbol{\eta}_x = (\eta_{rx}, \dots, \eta_{r-1,x})^T, \boldsymbol{\xi}_x = (\xi_{1x}, \dots, \xi_{c-1,x})^T.$$

La restricción de  $Z_1$  y  $Z_2$  a tomar valores enteros no reduce la generalidad del modelo,

ya que la función de probabilidad acumulada bivalente de  $\mathbf{Z}(\mathbf{x})$  dependerá de  $Z_1(\mathbf{x})$  y  $Z_2(\mathbf{x})$  únicamente a través de  $\boldsymbol{\eta}_x$  y  $\boldsymbol{\xi}_x$ .

La función de probabilidad acumulada de  $\mathbf{Z}(\mathbf{x})$  es

$$F_{ij}(\mathbf{x}; \boldsymbol{\beta}_x) = \Pr(Z_1(\mathbf{x}) \leq i, Z_2(\mathbf{x}) \leq j) \quad (2.4)$$

con  $\boldsymbol{\beta}_x = (\boldsymbol{\eta}_x, \boldsymbol{\xi}_x, \boldsymbol{\psi}_x)$ , donde  $\boldsymbol{\psi}_x$  es una matriz  $(r-1) \times (c-1)$  de “cross-ratios” globales definida más adelante. Dividimos el espacio de los posibles valores de  $\mathbf{Z}$  en cuatro cuadrantes,

$$\{Z_1 \leq i, Z_2 \leq j\}, \quad \{Z_1 \leq i, Z_2 > j\}, \quad \{Z_1 > i, Z_2 \leq j\}, \quad \{Z_1 > i, Z_2 > j\}.$$

Esta doble dicotomía ocurrirá en el punto de corte bivalente  $(i, j)$ . La tabla  $r \times c$  de probabilidades  $\pi_{ij|x} = \Pr(Z_1 = i, Z_2 = j)$  se comprime en una tabla  $2 \times 2$ .

Se define el “cross-ratio” global (*CRG*) en el punto  $(i, j)$  como

$$\begin{aligned} \psi_{ijx} &= \frac{\Pr[Z_1 \leq i, Z_2 \leq j] \Pr[Z_1 > i, Z_2 > j]}{\Pr[Z_1 > i, Z_2 \leq j] \Pr[Z_1 \leq i, Z_2 > j]} \\ &= \frac{F_{ij}(\mathbf{x}, \boldsymbol{\beta}_x) [1 - \eta_{ix} - \xi_{jx} + F_{ij}(\mathbf{x}, \boldsymbol{\beta}_x)]}{[\xi_{jx} - F_{ij}(\mathbf{x}, \boldsymbol{\beta}_x)] [\eta_{ix} - F_{ij}(\mathbf{x}, \boldsymbol{\beta}_x)]}. \end{aligned} \quad (2.5)$$

A partir de (2.5),  $F_{ij}(\mathbf{x}, \boldsymbol{\beta}_x)$  se puede obtener en términos de  $\boldsymbol{\eta}_x$ ,  $\boldsymbol{\xi}_x$ ,  $\boldsymbol{\psi}_x$

$$F_{ij}(\mathbf{x}; \boldsymbol{\eta}_x, \boldsymbol{\xi}_x, \boldsymbol{\psi}_x) = \begin{cases} \frac{1}{2} (\psi_{ijx} - 1)^{-1} [1 + (\eta_{ix} + \xi_{jx}) (\psi_{ijx} - 1) - S(\eta_{ix}, \xi_{jx}, \psi_{ijx})] & \text{si } \psi_{ijx} \neq 1 \\ \eta_{ix} \xi_{jx} & \text{si } \psi_{ijx} = 1 \end{cases} \quad (2.6)$$

con  $i = 1, \dots, r-1$ ;  $j = 1, \dots, c-1$  y donde

$$S(\eta, \xi, \psi) = \left\{ [1 + (\eta + \xi)(\psi - 1)]^2 + 4\psi(1 - \psi)\eta\xi \right\}^{1/2}. \quad (2.7)$$

La distribución definida en (2.6) se le conoce por **Modelo “Cross-Ratio” Global (CRG)** para  $\mathbf{Z}(\mathbf{x})$ .

$F$  es una función de distribución acumulada siempre que

$$F_{ijx} - F_{i-1,j,x} - F_{i,j-1,x} + F_{i-1,j-1,x} \geq 0,$$

donde

$$F_{ijx} = F_{ij}(\mathbf{x}; \boldsymbol{\beta}_x), \quad i = 1, \dots, r, \quad j = 1, \dots, c.$$

Esto requiere que las probabilidades marginales acumuladas satisfagan

$$0 < \eta_{1x} < \dots < \eta_{r-1,x} < 1, \quad 0 < \xi_{1x} < \dots < \xi_{c-1,x} < 1$$

y que las  $\psi_{ijx}$  para puntos adyacentes no sean muy diferentes. Cuando  $\psi_{ijx} = \psi(x)$ , es decir, no depende de los puntos de corte, entonces  $\psi(x)$  no está restringida, ver Mardia [65]. Si  $\eta_{ix} = 1$  entonces  $F_{ijx} = \xi_{jx}$ , cualquiera que sea el valor de  $\psi_{ijx}$ .  $F_{ijx}$  depende de  $\mathbf{x}$  a través de  $\eta_{ix}$ ,  $\xi_{jx}$  y por  $\psi_{ijx}$ .

Para un  $\mathbf{x}$  dado, el modelo CRG general de (2.6) es saturado, tiene  $r \times c - 1$  parámetros, igual al n° de grados de libertad en la tabla. Para realizar las estimaciones, la dimensión del espacio paramétrico se reduce poniendo restricciones tanto en los parámetros CRG como en los marginales.

### 2.2.1. Parametrización Lineal para Modelos CRG

Con objeto de simplificar la matriz general  $\boldsymbol{\psi}_x$  de los parámetros CRG  $\psi_{ijx}$ , el CRG puede depender de  $\mathbf{x}$  o de los puntos de corte. Ya que  $\psi_{ijx}$  está definido en el intervalo  $(0, \infty)$ , es preferible modelizar  $\log(\psi_{ijx})$ . Una forma general es

$$\log(\psi_{ijx}) = \Delta + \alpha_{ia} + \beta_{ja} + \delta_{ij} - \boldsymbol{\gamma}^T \mathbf{x} \quad (2.8)$$

donde  $i = 1, \dots, r-1$ ,  $j = 1, \dots, c-1$  y  $a$  es el parámetro de asociación; con restricciones de unicidad simples  $\alpha_{r-1,a} = \beta_{c-1,a} = 0$ ;  $\delta_{i,c-1} = 0$ ,  $i = 1, \dots, r-1$ ,  $\delta_{r-1,j} = 0$ ,  $j = 1, \dots, c-1$ . Entonces  $\boldsymbol{\gamma} = \mathbf{0}$  indica que la asociación no depende de  $\mathbf{x}$ ; los  $\delta_{ij}$  cero indica que no hay interacción entre las dos características observadas, etc. Cuando  $\psi_{ijx}$  no depende de los puntos de corte, tenemos un modelo CRG constante, ya que dado  $\mathbf{x}$  el CRG es constante para todos los puntos de corte. Este modelo fue estudiado por Pearson y Heron [88] y por Plackett [89] para respuesta binaria con distribuciones marginales,

continuas o discretas, especificadas. Modelos para respuestas trivariantes o de mayor dimensión se pueden definir en términos de una medida de asociación análoga al CRG, ver Anscombe [9] (pp.310-311).

Las probabilidades marginales acumuladas se ajustan utilizando los modelos lineales generalizados propuestos por McCullagh [67]. Dadas las funciones enlace  $g_1(\cdot)$ ,  $g_2(\cdot)$ , funciones de probabilidad monótonas diferenciables, se considera

$$\begin{aligned} g_1(\eta_{ix}) &= \alpha_{1i} - \mathbf{x}^T \boldsymbol{\beta}_1 & i = 1, \dots, r-1 \\ g_2(\xi_{jx}) &= \alpha_{2j} - \mathbf{x}^T \boldsymbol{\beta}_2 & j = 1, \dots, c-1. \end{aligned} \quad (2.9)$$

donde  $\boldsymbol{\beta}_1$  y  $\boldsymbol{\beta}_2$  son parámetros  $p$ -dimensionales que no dependen de los puntos de corte. Para  $g_q(s) = \log[s/(1-s)]$ ,  $q = 1, 2$ , (2.9) se convierten en los modelos de regresión lineal logística

$$\begin{aligned} \eta_{ix} &= \exp(\alpha_{1i} - \mathbf{x}^T \boldsymbol{\beta}_1) / [1 + \exp(\alpha_{1i} - \mathbf{x}^T \boldsymbol{\beta}_1)] & i = 1, \dots, r-1 \\ \xi_{jx} &= \exp(\alpha_{2j} - \mathbf{x}^T \boldsymbol{\beta}_2) / [1 + \exp(\alpha_{2j} - \mathbf{x}^T \boldsymbol{\beta}_2)] & j = 1, \dots, c-1. \end{aligned} \quad (2.10)$$

Eligiendo las funciones enlace  $g_1$ ,  $g_2$  para modelar las marginales, como en (2.9) y (2.8) para los parámetros del CRG, junto con (2.6) define una familia paramétrica de modelos CRG.

El CRG puede ser interpretado como un “odd ratio” de sucesos condicionales, los “odds” en  $\Pr[Z_1 \leq i \mid Z_2 \leq j]$  frente a los “odds”  $\Pr[Z_1 \leq i \mid Z_2 > j]$ , donde los “odds” en el suceso  $E \equiv \Pr(E) / [1 - \Pr(E)]$ . Si  $j = c$  entonces los segundos “odds ratios” son degenerados, ya que  $\Pr[Z_2 > c] = 0$ , el primer “odds” es  $[Z_i(\mathbf{x}) \leq i]$ . Las probabilidades marginales acumuladas son modeladas de forma similar al  $\log(CRG)$  usando logits.

### Algunas Propiedades del Modelo de Dale Bivariante

1. La flexibilidad con la que se modeliza la estructura marginal. La probabilidad marginal acumulada puede ser ajustada en el marco del modelo lineal generalizado.
2. Los parámetros marginales son ortogonales en los parámetros de asociación en el sentido en que los elementos correspondientes en la matriz de covarianzas esperada son idénticamente nulos (Palmgren [77]).



Los modelos CRG no necesitan valores de las marginales: son invariantes bajo cualquier transformación monótona de las variables respuesta marginales, y están definidos incluso cuando conocemos sólo un orden de las categorías marginales. Si algunas categorías marginales adyacentes se combinan, los modelos CRG para la nueva tabla tienen menos parámetros. Sin embargo, los parámetros que quedan tienen la misma interpretación que en el modelo original. Esto no ocurre para modelos basados en asociación local (Goodman [47]).

### 2.2.2. Ajuste Modelo de Dale por máxima verosimilitud

Un método sencillo de estimación que tiene buenas propiedades asintóticas en nuestra situación es la máxima verosimilitud, Cox et al. [30]. La estimación de los parámetros por máxima verosimilitud para la mayoría de los modelos de la sección anterior no tienen una fórmula explícita por lo que se tiene que utilizar una aproximación numérica.

Consideramos  $n_k$  observaciones independientes distribuidas de acuerdo a la familia de modelos CRG con variables explicativas  $\mathbf{x}_k$ , marginales ajustadas por regresión logística (2.10), y parámetros CRG como en (2.8). Sea  $Y_{ijk}$  el número de observaciones en la celda  $(i, j)$ , con  $\mathbf{Y}_k$  denotando la tabla completa  $r \times c$ . Entonces la distribución de  $\mathbf{Y}_k$  es multinomial con  $n_k$  observaciones y  $r \times c$  celdas. Ajustamos este modelo CRG a una secuencia de  $m$  tablas independientes  $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ , dependientes de una matriz de variables explicativas  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$ . La función de log-verosimilitud para  $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ , excepto por un término sumado independiente de  $\boldsymbol{\theta}$  es

$$\ell(\boldsymbol{\theta}; \mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{X}) = \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c y_{ijk} \log [\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})]$$

donde

$$\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta}) = \Pr[Z_1 = i, Z_2 = j \mid \boldsymbol{\theta}] \quad (2.11)$$

$$\boldsymbol{\theta} = (\alpha_1, \beta_1, \alpha_2, \beta_2, \Delta, \alpha_a, \beta_a, \delta, \gamma)$$

y las componentes de  $\boldsymbol{\theta}$  están definidos en (2.8) y (2.9). Dale [36] maximiza  $\ell$  utilizando un método cuasi-Newton de las librerías NAG (1978) para ello calcula algebraicamente el valor de la función de log-verosimilitud y sus primeras derivadas. La matriz de las segundas derivadas parciales es obtenida numéricamente con el procedimiento de ma-

ximización: menos la inversa en el máximo proporciona la estimación de la matriz de covarianzas, o lo que es lo mismo, la matriz de información de Fisher observada.

Para empezar la rutina de maximización, se necesita un valor inicial de los parámetros. Los punto iniciales que propone este autor son los siguientes: Para los parámetros de los puntos de corte  $\alpha_1$  y  $\alpha_2$  en el modelo logístico marginal de (2.10), aplica una aproximación parecida a la que propone Cox [29] para Binomiales, tratando las filas y las columnas marginales por separado. Los parámetros  $\beta_1$  y  $\beta_2$  son inicializados a cero. Para  $k = 1, \dots, m$  y para cada punto de corte  $(i, j)$  una estimación de los individuos en el primer cuadrante es

$$A_{ij/k} = \frac{1}{2} + \sum_{p \leq i} \sum_{q \leq j} \mathbf{Y}_{pqk} \quad (2.12)$$

en el segundo cuadrante

$$B_{ij/k} = \frac{1}{2} + \sum_{p \leq i} \sum_{q > j} \mathbf{Y}_{pqk} \quad (2.13)$$

en el tercer cuadrante

$$C_{ij/k} = \frac{1}{2} + \sum_{p > i} \sum_{q \leq j} \mathbf{Y}_{pqk} \quad (2.14)$$

y en el cuarto cuadrante

$$D_{ij/k} = \frac{1}{2} + \sum_{p > i} \sum_{q > j} \mathbf{Y}_{pqk}. \quad (2.15)$$

Sea

$$w_{ij/k} = \left( A_{ij/k}^{-1} + \dots + D_{ij/k}^{-1} \right)^{-1}$$

y

$$\tilde{\Delta}_{ij/k} = \log \left[ (A_{ij/k} D_{ij/k}) / (B_{ij/k} C_{ij/k}) \right]. \quad (2.16)$$

Se puede ver en Anscombe [9] que  $\tilde{\Delta}_{ij/k}$  tiene de media  $\log(\psi_{ijx})$  y varianza  $E(w_{ij/k}^{-1})$  aproximadamente para  $n_k$  grande. De este modo, un estimador aproximadamente insesgado de  $\Delta_{ij}$  es

$$\Delta_{ij} = \left( \sum_k w_{ij/k} \tilde{\Delta}_{ij/k} \right) / \left( \sum_k w_{ij/k} \right).$$

Una extensión de Anscombe [9], nos sirve para calcular la estimación inicial de  $\Delta$  :

$$\tilde{\Delta} = \left( \sum_k n_k \right)^{-1} \sum_k n_k W_k^{-1} \left[ \sum_i \sum_j w_{ij/k} \tilde{\Delta}_{ij/k} \right]$$

donde  $W_k = \sum_i \sum_j w_{ij/k}$ ,  $k = 1, \dots, m$ .

## 2.3. Ajuste del Modelo de Dale por mínima $\phi$ -divergencia

Es interesante observar que el logaritmo de la logverosimilitud salvo constantes se puede escribir alternativamente como

$$\begin{aligned}
\ell(\boldsymbol{\theta}; \mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{X}) &= \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c y_{ijk} \log [\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})] \\
&= -n \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{y_{ijk}}{n} \left[ \log [\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})]^{-1} \right] \\
&\quad + n \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{y_{ijk}}{n} \log \frac{y_{ijk}}{n_k} - n \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{y_{ijk}}{n} \log \frac{y_{ijk}}{n_k} \\
&= Cte - n \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{y_{ijk}}{n} \log \left[ \frac{\frac{y_{ijk}}{n_k}}{\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \right] \\
&= Cte - n \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{y_{ijk}}{n} \log \left[ \frac{\frac{y_{ijk}}{n}}{\frac{\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta}) n_k}{n}} \right] \\
&= Cte - n D_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))
\end{aligned}$$

siendo  $n = \sum_{k=1}^m n_k$  y  $\hat{\mathbf{p}}$  y  $\mathbf{p}(\boldsymbol{\theta})$  vectores de probabilidad de dimensión  $r \times c \times m$  definidos como

$$\begin{aligned}
\hat{\mathbf{p}} &= \left( \frac{y_{111}}{n}, \dots, \frac{y_{112}}{n}, \dots, \frac{y_{rcm}}{n} \right)^T \\
\mathbf{p}(\boldsymbol{\theta}) &= \left( \frac{n_1 \pi_{11}(\mathbf{x}_1; \boldsymbol{\theta})}{n}, \dots, \frac{n_2 \pi_{11}(\mathbf{x}_2; \boldsymbol{\theta})}{n}, \dots, \frac{n_m \pi_{rc}(\mathbf{x}_m; \boldsymbol{\theta})}{n} \right)^T.
\end{aligned}$$

De esta forma, el estimador de máxima verosimilitud (EMV) de  $\boldsymbol{\theta}$ ,  $\hat{\boldsymbol{\theta}}$ , se puede definir de forma equivalente como

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} D_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$$

donde

$$\boldsymbol{\Theta} = \{\boldsymbol{\theta} = (\alpha_1, \beta_1, \alpha_2, \beta_2, \boldsymbol{\Delta}, \alpha_a, \beta_a, \boldsymbol{\delta}, \boldsymbol{\gamma}) \in \mathbb{R}^p\}$$

siendo  $p$  el número de parámetros desconocidos.

La medida de divergencia de Kullback es un caso particular de la medida de  $\phi$ -divergencia definida por Csiszár [34]

$$\begin{aligned} D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) &= \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \phi \left( \frac{\frac{y_{ijk}}{n}}{\frac{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{n}} \right) \\ &= \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \phi \left( \frac{y_{ijk}}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \right) \end{aligned} \quad (2.17)$$

siendo  $\phi \in \Phi^*$  la clase de todas las funciones convexas  $\phi(x)$ ,  $x > 0$ , tales que en  $x = 1$ ,  $\phi(1) = \phi'(1) = 0$ ,  $\phi''(1) > 0$ , y en  $x = 0$ ,  $0\phi(0/0) = 0$  y  $0\phi(p/0) = p \lim_{u \rightarrow \infty} \frac{\phi(u)}{u}$ .

Por tanto, una extensión natural del EMV viene dada por el estimador de mínima  $\phi$ -divergencia.

**Definición 2.1** Sean  $\mathbf{Y}_k$ ,  $k = 1, \dots, m$ , variables aleatorias independientes con distribución multinomial de parámetros  $(n_k, \boldsymbol{\pi}(\mathbf{x}_k; \boldsymbol{\theta}))$  con  $\boldsymbol{\pi}(\mathbf{x}_k; \boldsymbol{\theta}) = \left( \frac{n_k \pi_{11}(\mathbf{x}_k; \boldsymbol{\theta})}{n}, \dots, \frac{n_k \pi_{rc}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \right)$ . Se denomina estimador de mínima  $\phi$ -divergencia de  $\boldsymbol{\theta}$  para el modelo **CRG** dado en (2.8) y (2.9) al vector

$$\hat{\boldsymbol{\theta}}_\phi = \arg \min_{\boldsymbol{\theta} \in \Theta} D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})).$$

**Observación 2.1** Para  $\phi(x) = x \log x - x + 1$  se obtiene el EMV.

**Proposición 2.1** La matriz de información de Fisher para el modelo CRG dado en (2.8) viene dada por

$$\begin{aligned} \mathbf{I}_{F,n}(\boldsymbol{\theta}) &= (i_{(s,l)}(\boldsymbol{\theta}))_{\substack{s=1,\dots,p \\ l=1,\dots,p}} \\ &= \left( \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{n_k}{n \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l} \right)_{\substack{s=1,\dots,p \\ l=1,\dots,p}}. \end{aligned} \quad (2.18)$$

**Demostración.** Al ser

$$\ell(\boldsymbol{\theta}; \mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{X}) = \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c y_{ijk} \log [\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})]$$

se tiene

$$\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{X})}{\partial \theta_s} = \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{y_{ijk}}{\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s}$$

por tanto

$$\begin{aligned} i_{(s, l)}(\boldsymbol{\theta}) &= \frac{1}{n} E \left[ \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X})}{\partial \theta_s} \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X})}{\partial \theta_l} \right] \\ &= \frac{1}{n} E \left[ \sum_{k_1, k_2=1}^m \sum_{i_1, i_2=1}^r \sum_{j_1, j_2=1}^c \frac{y_{i_1 j_1 k_1}}{\pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})} \frac{\partial \pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})}{\partial \theta_s} \right. \\ &\quad \left. \times \frac{y_{i_2 j_2 k_2}}{\pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})} \frac{\partial \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})}{\partial \theta_l} \right] \\ &= \frac{1}{n} E \left[ \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \left( \frac{y_{ijk}}{\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \right)^2 \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l} \right] \\ &\quad + \frac{1}{n} E \left[ \sum_{\substack{k_1, k_2=1 \\ \text{excepto } k_1=k_2, i_1=i_2, j_1=j_2}}^m \sum_{i_1, i_2=1}^r \sum_{j_1, j_2=1}^c \frac{y_{i_1 j_1 k_1}}{\pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})} \frac{\partial \pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})}{\partial \theta_s} \frac{y_{i_2 j_2 k_2}}{\pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})} \frac{\partial \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})}{\partial \theta_l} \right] \end{aligned}$$

Pero

$$E[\mathbf{Y}_{i_1 j_1 k_1} \mathbf{Y}_{i_2 j_2 k_2}] = \begin{cases} \left[ n \frac{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \left( 1 - \frac{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \right) + n^2 \left( \frac{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \right)^2 \right] & \begin{matrix} i_1 = i_2 = i \\ j_1 = j_2 = j \\ k_1 = k_2 = k \end{matrix} \\ \left[ -n \frac{n_{k_1} \pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})}{n} \frac{n_{k_2} \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})}{n} + n^2 \frac{n_{k_1} \pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})}{n} \frac{n_{k_2} \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})}{n} \right] & \text{en caso contrario} \end{cases}$$

Entonces tenemos

$$\begin{aligned}
 i_{(s, l)}(\boldsymbol{\theta}) &= \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{\left[ \frac{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \left( 1 - \frac{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \right) + n \left( \frac{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \right)^2 \right]}{(\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta}))^2} \\
 &\quad \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l} \\
 &+ \sum_{\substack{k_1, k_2=1 \\ \text{excepto } k_1=k_2, i_1=i_2, j_1=j_2}}^m \sum_{i_1, i_2=1}^r \sum_{j_1, j_2=1}^c \left[ \frac{-\frac{n_{k_1} \pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})}{n} \frac{n_{k_2} \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})}{n}}{\pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta}) \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})} \right. \\
 &\quad \left. + \frac{n \frac{n_{k_1} \pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})}{n} \frac{n_{k_2} \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})}{n}}{\pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta}) \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})} \right] \times \frac{\partial \pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})}{\partial \theta_l} \\
 &= \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \left[ \frac{n_k}{n \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} - \frac{n_k^2}{n^2} + \frac{n_k^2}{n} \right] \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l} \\
 &+ \sum_{\substack{k_1, k_2=1 \\ \text{excepto } k_1=k_2, i_1=i_2, j_1=j_2}}^m \sum_{i_1, i_2=1}^r \sum_{j_1, j_2=1}^c \left[ -\frac{n_{k_1} n_{k_2}}{n^2} + \frac{n_{k_1} n_{k_2}}{n} \right] \frac{\partial \pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})}{\partial \theta_l} \\
 &= \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{n}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l} \\
 &+ \sum_{k_1, k_2=1}^m \sum_{i_1, i_2=1}^r \sum_{j_1, j_2=1}^c (n-1) \frac{n_{k_1}}{n} \frac{\partial \pi_{i_1 j_1}(\mathbf{x}_{k_1}; \boldsymbol{\theta})}{\partial \theta_s} \frac{n_{k_2}}{n} \frac{\partial \pi_{i_2 j_2}(\mathbf{x}_{k_2}; \boldsymbol{\theta})}{\partial \theta_l} \\
 &= \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{n_k}{n \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l}
 \end{aligned}$$

■

**Observación 2.2** La estimación de los parámetros por mínima  $\phi$ -divergencia no tiene una fórmula explícita por lo que se tiene que utilizar una aproximación numérica al igual que para obtener el EMV.

### 2.3.1. Comportamiento asintótico del estimador de mínima $\phi$ -divergencia

En este apartado se estudian las propiedades asintóticas del estimador de mínima  $\phi$ -divergencia definido en el apartado anterior.

El primer objetivo es el de encontrar la descomposición BAN (Best Asymptotically Normal) del estimador de mínima  $\phi$ -divergencia. Existen diversas formas de llegar a esta descomposición, nosotros haremos uso del Teorema de la Función Implícita.

**Teorema 2.1 (Teorema de la Función Implícita)** *Sea*

$$\mathbf{F} = (F_1, \dots, F_p) : \mathbb{R}^{m+p} \rightarrow \mathbb{R}^p$$

*una función diferenciable con continuidad en un conjunto abierto  $D \subset \mathbb{R}^{M+p}$  de forma que este conjunto contiene al punto*

$$\left( \mathbf{x}_0 = (x_1^0, \dots, x_m^0)^T, \mathbf{y}_0 = (y_1^0, \dots, y_p^0)^T \right)$$

*verificándose para este punto que  $\mathbf{F}(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$ . Supongamos que la matriz*

$$\mathbf{J}_F = \left( \frac{\partial F_i}{\partial y_j} \right)_{\substack{i=1, \dots, p \\ j=1, \dots, p}}$$

*es no singular para  $(\mathbf{x}_0, \mathbf{y}_0)$ . Entonces existe un entorno  $\mathbf{U}$  de  $(\mathbf{x}_0, \mathbf{y}_0)$  de forma que la matriz  $\mathbf{J}_F$  es no singular para todos los puntos  $(\mathbf{x}, \mathbf{y}) \in \mathbf{U}$ , un conjunto abierto  $A \subset \mathbb{R}^M$  conteniendo  $\mathbf{x}_0$  y una única función diferenciable con continuidad  $\mathbf{g} = (g_1, \dots, g_p) : A \rightarrow \mathbb{R}^p$  de forma que*

$$\{(\mathbf{x}, \mathbf{y}) \in \mathbf{U} : \mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{0}\} = \{(\mathbf{x}, \mathbf{g}(\mathbf{x})) : \mathbf{x} \in A\}.$$

**Demostración.** La demostración puede verse por ejemplo en la página 148 del libro de Flemming [44]. ■

El siguiente Teorema presenta la descomposición BAN del estimador de mínima  $\phi$ -divergencia.

**Teorema 2.2** *Sean  $\mathbf{Y}_k$ ,  $k = 1, \dots, m$ , variables aleatorias independientes con distribución multinomial de parámetros  $(n_k, \boldsymbol{\pi}(\mathbf{x}_k; \boldsymbol{\theta}))$  con*

$$\boldsymbol{\pi}(\mathbf{x}_k; \boldsymbol{\theta}) = \left( \frac{n_k \pi_{11}(\mathbf{x}_k; \boldsymbol{\theta})}{n}, \dots, \frac{n_k \pi_{rc}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \right).$$

Sea

$$\Delta_{n\_celdas} = \left\{ \mathbf{q} = (q_1, \dots, q_{n\_celdas})^T : \sum_{i=1}^{n\_celdas} q_i = 1, q_i > 0, i = 1, \dots, n\_celdas \right\},$$

con  $n\_celdas = r \times c \times m$  y supongamos que  $\mathbf{p} : \Theta \rightarrow \Delta_{n\_celdas}$  tiene derivadas parciales segundas continuas en un entorno del verdadero valor del parámetro  $\theta^0$  y sea  $\phi(t) \in \Phi^*$  dos veces continuamente diferenciable para  $t > 0$  con  $\mathbf{I}_{F,n}$ , definida positiva en el verdadero vector de parámetro  $\theta^0$ . El estimador de mínima  $\phi$ -divergencia de  $\theta$  es único en un entorno de  $\theta^0$  para el modelo CRG dado en (2.8) y verifica que

$$\begin{aligned} \hat{\theta}_\phi = & \theta^0 + \mathbf{I}_{F,n}(\theta^0)^{-1} \left( \frac{\partial \pi(\mathbf{X}; \theta)}{\partial \theta} \right)_{\theta=\theta^0}^T \text{Diag} \left( \pi(\mathbf{X}; \theta^0)^{-1} \right) (\hat{\mathbf{p}} - \mathbf{p}(\theta^0)) \\ & + \|\hat{\mathbf{p}} - \mathbf{p}(\theta^0)\| \alpha_1(\hat{\mathbf{p}}; \hat{\mathbf{p}} - \mathbf{p}(\theta^0)) \end{aligned} \quad (2.19)$$

donde  $\mathbf{I}_{F,n}(\theta^0)$  definido en (2.8),  $\pi(\mathbf{X}; \theta) = (\pi(\mathbf{x}_1; \theta), \dots, \pi(\mathbf{x}_m; \theta))$  y

$$\alpha_1 : \mathbb{R}^{n\_celdas \times n\_celdas} \rightarrow \mathbb{R}^p \text{ verifica que } \alpha_1(\mathbf{p}; \mathbf{p} - \mathbf{p}(\theta^0)) \rightarrow \mathbf{0} \text{ cuando } \mathbf{p} \rightarrow \mathbf{p}(\theta^0). \quad (2.20)$$

**Demostración.** Sea  $l^{n\_celdas}$  el interior del cubo unitario de dimensión  $n\_celdas$  con  $\Delta_{n\_celdas} \subset l^{n\_celdas}$ .  $\mathbf{W}$  un entorno de  $\theta^0$  en el cual la función  $\mathbf{p} : \Theta \rightarrow \Delta_{n\_celdas}$  tiene derivadas parciales segundas continuas y  $p$  el número de parámetros desconocidos. Se define la función vectorial

$$\mathbf{F} = (F_1, \dots, F_p)$$

de  $l^{n\_celdas} \times \mathbf{W}$  en  $\mathbf{W}$  de forma que

$$F_s(\mathbf{p}, \theta) = \frac{\partial D_\phi(\mathbf{p}, \mathbf{p}(\theta))}{\partial \theta_s}, \quad s = 1, \dots, p,$$

con  $\mathbf{p} = (p_{111}, \dots, p_{11c}, \dots, p_{m11}, \dots, p_{mrc}) \in l^{n\_celdas}$ . La función  $\mathbf{F}$  es una función diferenciable con continuidad en el conjunto abierto  $l^{n\_celdas} \times \mathbf{W}$ . Además

$$F_s(\mathbf{p}(\theta^0), \theta^0) = 0, \quad s = 1, \dots, p.$$



Puesto que

$$F_s(\mathbf{p}, \boldsymbol{\theta}) = \frac{\partial D_\phi(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_s} = \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \left[ \frac{n_k}{n} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \phi\left(\frac{np_{ijk}}{n_k \pi_{ijk}}\right) - \frac{p_{kij}}{\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \phi'\left(\frac{np_{ijk}}{n_k \pi_{ijk}}\right) \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \right] \quad (2.21)$$

se tiene

$$\begin{aligned} \frac{\partial F_s(\mathbf{p}, \boldsymbol{\theta})}{\partial \theta_l} = & - \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \phi'\left(\frac{np_{kij}}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}\right) \frac{p_{kij}}{(\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta}))^2} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l} \\ & + \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \phi''\left(\frac{np_{kij}}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}\right) \frac{p_{kij}}{n_k (\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta}))^2} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l} \frac{p_{kij}}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \\ & + \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \phi'\left(\frac{np_{kij}}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}\right) \frac{p_{kij}}{(\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta}))^2} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l} \\ & + \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{\partial^2 \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l \partial \theta_s} \left( \frac{n_k}{n} \phi\left(\frac{np_{kij}}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}\right) - \phi'\left(\frac{np_{kij}}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}\right) \frac{p_{kij}}{\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \right), \end{aligned}$$

y

$$\begin{aligned} \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}^0} &= \left( \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \\ &= \left( \left( \frac{\partial}{\partial \theta_s} \left( \frac{\partial D_\phi(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_l} \right) \right)_{\substack{s=1, \dots, p \\ l=1, \dots, p}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \\ &= \phi''(1) \left( \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \frac{n_k}{n \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta}^0)} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})}{\partial \theta_l} \right)_{\substack{s=1, \dots, p \\ l=1, \dots, p}} \\ &= \phi''(1) \mathbf{I}_{F,n}(\boldsymbol{\theta}^0). \end{aligned}$$

Esta matriz es definida positiva en  $\boldsymbol{\theta}^0$  ya que por hipótesis  $\phi''(1) > 0$  y la matriz de Información de Fisher para el modelo es definida positiva. Aplicando el Teorema de la Función Implícita existe un entorno  $\mathbf{U}$  de  $(\mathbf{p}(\boldsymbol{\theta}^0), \boldsymbol{\theta}^0)$  de forma que la matriz

$$\mathbf{J}_{\mathbf{F}} = \left( \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}} \right)$$

es definida positiva para todo  $(\mathbf{p}, \boldsymbol{\theta}) \in \mathbf{U}$  ya que  $\mathbf{F}$  es continuamente diferenciable. También existe un abierto  $A \subset l^{n-celdas}$  y una única función diferenciable con continuidad

$$\tilde{\boldsymbol{\theta}} : A \subset l^{n-celdas} \rightarrow \mathbb{R}^p$$

de forma que  $\mathbf{p}(\boldsymbol{\theta}^0) \in A$  y

$$\{(\mathbf{p}, \boldsymbol{\theta}) \in \mathbf{U} : \mathbf{F}(\mathbf{p}, \boldsymbol{\theta}) = 0\} = \left\{ \left( \mathbf{p}, \tilde{\boldsymbol{\theta}}(\mathbf{p}) \right) : \mathbf{p} \in A \right\}. \quad (2.22)$$

Es claro que en  $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\theta}^0))$  la función

$$D_\phi(\mathbf{p}(\boldsymbol{\theta}^0), \mathbf{p}(\boldsymbol{\theta}))$$

alcanza un mínimo ya que como  $\mathbf{p}(\boldsymbol{\theta}^0) \in A$  se tiene que

$$\mathbf{F}(\mathbf{p}(\boldsymbol{\theta}^0), \tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\theta}^0))) = \frac{\partial D_\phi(\mathbf{p}(\boldsymbol{\theta}^0), \mathbf{p}(\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\theta}^0))))}{\partial \boldsymbol{\theta}} = 0$$

y teniendo en cuenta (2.22),

$$(\mathbf{p}(\boldsymbol{\theta}^0), \tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\theta}^0))) \in \mathbf{U}$$

por lo cual  $\mathbf{J}_{\mathbf{F}}$  es definida positiva en

$$(\mathbf{p}(\boldsymbol{\theta}^0), \tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\theta}^0))).$$

Es decir,

$$D_\phi(\mathbf{p}(\boldsymbol{\theta}^0), \mathbf{p}(\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\theta}^0)))) = \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} D_\phi(\mathbf{p}(\boldsymbol{\theta}^0), \mathbf{p}(\boldsymbol{\theta}))$$

y por las propiedades de la  $\phi$ -divergencia  $\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\theta}^0)) = \boldsymbol{\theta}^0$ .

Sabemos que

$$\hat{\mathbf{p}} \xrightarrow[n \rightarrow \infty]{c.s.} \mathbf{p}(\boldsymbol{\theta}^0),$$

por lo que  $\hat{\mathbf{p}} \in A$ . Entonces  $(\hat{\mathbf{p}}, \tilde{\boldsymbol{\theta}}(\hat{\mathbf{p}})) \in \mathbf{U}$  y como consecuencia  $\tilde{\boldsymbol{\theta}}(\hat{\mathbf{p}})$  es solución del sistema de ecuaciones

$$\frac{\partial D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}(\hat{\mathbf{p}})))}{\partial \theta_j} = 0, \quad j = 1, \dots, p.$$

Por lo tanto  $\tilde{\boldsymbol{\theta}}(\hat{\mathbf{p}})$  es el estimador de mínima  $\phi$ -divergencia que está definido de forma única en  $\mathbf{U}$  por la única función  $\tilde{\boldsymbol{\theta}}(\mathbf{p})$ .

El desarrollo de Taylor de  $\tilde{\boldsymbol{\theta}}(\mathbf{p})$  en  $\mathbf{p}(\boldsymbol{\theta}^0)$  viene dado por,

$$\tilde{\boldsymbol{\theta}}(\mathbf{p}) = \tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\theta}^0)) + \left( \frac{\partial \tilde{\boldsymbol{\theta}}(\mathbf{p})}{\partial \mathbf{p}} \right)_{\mathbf{p}=\mathbf{p}(\boldsymbol{\theta}^0)} (\mathbf{p} - \mathbf{p}(\boldsymbol{\theta}^0)) + \|\mathbf{p} - \mathbf{p}(\boldsymbol{\theta}^0)\| \boldsymbol{\alpha}_1(\mathbf{p}; \mathbf{p} - \mathbf{p}(\boldsymbol{\theta}^0)) \quad (2.23)$$

donde  $\boldsymbol{\alpha}_1$  viene definida en el enunciado.

Aplicando ahora la regla de la cadena se tiene,

$$\frac{\partial \mathbf{F}(\mathbf{p}, \tilde{\boldsymbol{\theta}}(\mathbf{p}))}{\partial \mathbf{p}} + \frac{\partial \mathbf{F}(\mathbf{p}, \tilde{\boldsymbol{\theta}}(\mathbf{p}))}{\partial \tilde{\boldsymbol{\theta}}(\mathbf{p})} \frac{\tilde{\boldsymbol{\theta}}(\mathbf{p})}{\partial \mathbf{p}} = 0$$

y para  $\mathbf{p} = \mathbf{p}(\boldsymbol{\theta}^0)$ ,

$$\frac{\partial \mathbf{F}}{\partial \mathbf{p}(\boldsymbol{\theta}^0)} + \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}^0} \frac{\partial \boldsymbol{\theta}^0}{\partial \mathbf{p}(\boldsymbol{\theta}^0)} = 0.$$

Como

$$\frac{\partial \mathbf{F}}{\partial \mathbf{p}(\boldsymbol{\theta}^0)} = -\phi''(1) \left( \frac{\partial \mathbf{p}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}^T \text{Diag}(\mathbf{p}(\boldsymbol{\theta}^0)^{-1}),$$

y además

$$\left( \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \phi''(1) \mathbf{I}_{F,n}(\boldsymbol{\theta}^0)$$

de donde se tiene que

$$\frac{\partial \boldsymbol{\theta}^0}{\partial \mathbf{p}(\boldsymbol{\theta}^0)} = \mathbf{I}_{F,n}(\boldsymbol{\theta}^0)^{-1} \left( \frac{\partial \mathbf{p}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}^T \text{Diag}(\mathbf{p}(\boldsymbol{\theta}^0)^{-1}).$$

Al ser  $\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\theta}^0)) = \boldsymbol{\theta}^0$ , por (2.23) se tiene

$$\begin{aligned} \tilde{\boldsymbol{\theta}}(\mathbf{p}) &= \boldsymbol{\theta}^0 + \mathbf{I}_{F,n}(\boldsymbol{\theta}^0)^{-1} \left( \frac{\partial \mathbf{p}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}^T \text{Diag}(\mathbf{p}(\boldsymbol{\theta}^0)^{-1}) \\ &\quad \times (\mathbf{p} - \mathbf{p}(\boldsymbol{\theta}^0)) + \|\mathbf{p} - \mathbf{p}(\boldsymbol{\theta}^0)\| \boldsymbol{\alpha}_1(\mathbf{p}; \mathbf{p} - \mathbf{p}(\boldsymbol{\theta}^0)) \end{aligned}$$

y para  $\mathbf{p} = \hat{\mathbf{p}}$  se tiene el resultado enunciado. ■

A partir de este teorema se obtiene la normalidad asintótica del estimador de mínima  $\phi$ -divergencia en el modelo CRG.

**Teorema 2.3** Sean  $\mathbf{Y}_k$ ,  $k = 1, \dots, m$ , variables aleatorias independientes con distribución multinomial de parámetros  $(n_k, \boldsymbol{\pi}(\mathbf{x}_k; \boldsymbol{\theta}))$  con  $\boldsymbol{\pi}(\mathbf{x}_k; \boldsymbol{\theta}) = \left( \frac{n_k \pi_{11}(\mathbf{x}_k; \boldsymbol{\theta})}{n}, \dots, \frac{n_k \pi_{rc}(\mathbf{x}_k; \boldsymbol{\theta})}{n} \right)$ . Supongamos que  $\mathbf{p} : \Theta \rightarrow \Delta_{n\_celdas}$  tiene derivadas parciales segundas continuas en un

entorno de  $\theta^0$  y sea  $\phi(t) \in \Phi$  dos veces continuamente diferenciable para  $t > 0$  con  $\mathbf{I}_{F,n}$  definida positiva en  $\theta^0$  y  $n_i \rightarrow \infty$ ,  $i = 1, \dots, m$  tal que  $n_i/n \rightarrow \lambda_i > 0$ ,  $i = 1, \dots, m$ , entonces se tiene que

$$\sqrt{n} \left( \hat{\theta}_\phi - \theta^0 \right) \xrightarrow[n \rightarrow \infty]{L} \mathcal{N} \left( \mathbf{0}, \mathbf{I}_{F,\lambda} (\theta^0)^{-1} \right)$$

donde  $\mathbf{I}_{F,\lambda} (\theta) = \lim_{n \rightarrow \infty} \mathbf{I}_{F,n} (\theta)$ .

**Demostración.** Aplicando el Teorema Central del Límite se tiene que

$$\sqrt{n} \left( \hat{\mathbf{p}} - \mathbf{p} (\theta^0) \right) \xrightarrow[n \rightarrow \infty]{L} \mathcal{N} \left( \mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{p}_\lambda (\theta^0)} \right),$$

donde

$$\boldsymbol{\Sigma}_{\mathbf{p}_\lambda (\theta^0)} = \text{Diag} \left( \left( \mathbf{p}_\lambda (\theta^0) - \mathbf{p}_\lambda (\theta^0) \mathbf{p}_\lambda (\theta^0)^T \right)_{i=1, \dots, m} \right)$$

siendo  $\mathbf{p}_\lambda (\beta) = \lim_{n \rightarrow \infty} \mathbf{p} (\beta)$ .

Por tanto, por el Teorema 2.2 y lo anterior se tiene

$$\sqrt{n} \left( \hat{\theta}_\phi - \theta^0 \right) \xrightarrow[n \rightarrow \infty]{L} \mathcal{N} \left( \mathbf{0}, \boldsymbol{\Sigma}_1 (\theta^0) \right),$$

donde  $\boldsymbol{\Sigma}_1 (\theta^0)$  viene dada por

$$\begin{aligned} \boldsymbol{\Sigma}_1 (\theta^0) = & \mathbf{I}_{F,\lambda} (\theta^0)^{-1} \mathbf{A} (\theta^0)^T \text{Diag} \left( \mathbf{p}_\lambda (\theta^0)^{-1/2} \right) \boldsymbol{\Sigma}_{\mathbf{p}_\lambda (\theta^0)} \\ & \times \text{Diag} \left( \mathbf{p}_\lambda (\theta^0)^{-1/2} \right) \mathbf{A} (\theta^0) \mathbf{I}_{F,\lambda} (\theta^0)^{-1} \end{aligned}$$

con

$$\mathbf{A} (\theta) = \text{Diag} \left( \mathbf{p}_\lambda (\theta)^{-1/2} \right) \left( \frac{\partial \mathbf{p}_\lambda (\theta)}{\partial \theta} \right).$$

Ahora bien, llamando

$$\mathbf{B} = \text{Diag} \left( \mathbf{p}_\lambda (\theta^0)^{-1/2} \right) \boldsymbol{\Sigma}_{\mathbf{p}_\lambda (\theta^0)} \text{Diag} \left( \mathbf{p}_\lambda (\theta^0)^{-1/2} \right),$$

se tiene

$$\begin{aligned} \mathbf{B} = & \text{Diag} \left( \mathbf{p}_\lambda (\theta^0)^{-1/2} \right) \left( \text{Diag} \left( \mathbf{p}_\lambda (\theta^0) \right) - \mathbf{p}_\lambda (\theta^0) \mathbf{p}_\lambda (\theta^0)^T \right) \text{Diag} \left( \mathbf{p}_\lambda (\theta^0)^{-1/2} \right) \\ = & \text{Diag} \left( \mathbf{p}_\lambda (\theta^0)^{-1/2} \right) \text{Diag} \left( \mathbf{p}_\lambda (\theta^0) \right) \text{Diag} \left( \mathbf{p}_\lambda (\theta^0)^{-1/2} \right) \\ - & \text{Diag} \left( \mathbf{p}_\lambda (\theta^0)^{-1/2} \right) \mathbf{p}_\lambda (\theta^0) \mathbf{p}_\lambda (\theta^0)^T \text{Diag} \left( \mathbf{p}_\lambda (\theta^0)^{-1/2} \right) \\ = & \mathbf{I} - \sqrt{\mathbf{p}_\lambda (\theta^0)} \sqrt{\mathbf{p}_\lambda (\theta^0)^T}. \end{aligned}$$

Con lo cual,

$$\begin{aligned}\Sigma_1(\theta^0) &= (\mathbf{A}(\theta^0)^T \mathbf{A}(\theta^0))^{-1} \\ &\quad - (\mathbf{A}(\theta^0)^T \mathbf{A}(\theta^0))^{-1} \mathbf{A}(\theta^0)^T \sqrt{\mathbf{p}_\lambda(\theta^0)} \sqrt{\mathbf{p}_\lambda(\theta^0)}^T \mathbf{A}(\theta^0) (\mathbf{A}(\theta^0)^T \mathbf{A}(\theta^0))^{-1}.\end{aligned}$$

Veamos que  $\sqrt{\mathbf{p}_\lambda(\theta^0)}^T \mathbf{A}(\theta^0) = \mathbf{0}_{1 \times p}$ .

Se tiene,

$$\mathbf{A}(\theta^0) = \begin{pmatrix} \sqrt{\frac{\lambda_1}{\pi_{11}(\mathbf{x}_1; \theta^0)}} \frac{\partial \pi_{11}(\mathbf{x}_1; \theta^0)}{\partial \theta_1} & \cdots & \sqrt{\frac{\lambda_1}{\pi_{11}(\mathbf{x}_1; \theta^0)}} \frac{\partial \pi_{11}(\mathbf{x}_1; \theta^0)}{\partial \theta_p} \\ \vdots & \cdots & \vdots \\ \sqrt{\frac{\lambda_m}{\pi_{rc}(\mathbf{x}_m; \theta^0)}} \frac{\partial \pi_{rc}(\mathbf{x}_m; \theta^0)}{\partial \theta_1} & \cdots & \sqrt{\frac{\lambda_m}{\pi_{rc}(\mathbf{x}_m; \theta^0)}} \frac{\partial \pi_{rc}(\mathbf{x}_m; \theta^0)}{\partial \theta_p} \end{pmatrix}$$

entonces

$$\begin{aligned}\sqrt{\mathbf{p}_\lambda(\theta^0)}^T \mathbf{A}(\theta^0) &= \left( \sqrt{\lambda_1 \pi_{11}(\mathbf{x}_1; \theta^0)}, \dots, \sqrt{\lambda_m \pi_{rc}(\mathbf{x}_m; \theta^0)} \right) \mathbf{A}(\theta^0) \\ &= \left( \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \lambda_k \frac{\partial \pi_{ij}(\mathbf{x}_k; \theta^0)}{\partial \theta_1}, \dots, \sum_{k=1}^m \sum_{i=1}^r \sum_{j=1}^c \lambda_k \frac{\partial \pi_{ij}(\mathbf{x}_k; \theta^0)}{\partial \theta_p} \right) \\ &= (0, \dots, 0).\end{aligned}$$

Finalmente

$$\Sigma_1(\theta^0) = (\mathbf{A}(\theta^0)^T \mathbf{A}(\theta^0))^{-1} = \mathbf{I}_{F, \lambda}(\theta^0)^{-1}.$$

■

**Observación 2.3** Como caso particular del Teorema anterior se obtiene la normalidad asintótica del EMV

$$\sqrt{n}(\hat{\theta} - \theta^0) \xrightarrow[n \rightarrow \infty]{L} \mathcal{N}(\mathbf{0}, \mathbf{I}_{F, \lambda}(\theta^0)^{-1})$$

**Observación 2.4** Por los Teoremas 2.2 y 2.3 el estimador de mínima  $\phi$ -divergencia es BAN (Best Asymptotically normal).

A continuación se obtiene la distribución asintótica del vector de probabilidades,  $\mathbf{p}(\hat{\boldsymbol{\theta}}_\phi)$ , que caracteriza al modelo CRG dado en (2.8) cuando se estiman los parámetros por mínima  $\phi$ -divergencia.

**Teorema 2.4** *Bajo las hipótesis del Teorema 2.3 se tiene que*

$$\sqrt{n} \left( \mathbf{p}(\hat{\boldsymbol{\theta}}_\phi) - \mathbf{p}(\boldsymbol{\theta}^0) \right) \xrightarrow[n \rightarrow \infty]{L} \mathcal{N}(\mathbf{0}, \Sigma_2(\boldsymbol{\theta}^0))$$

donde

$$\Sigma_2(\boldsymbol{\theta}^0) = \mathbf{A}(\boldsymbol{\theta}^0)^T \mathbf{I}_{F,\lambda}(\boldsymbol{\theta}^0)^{-1} \mathbf{A}(\boldsymbol{\theta}^0)$$

con  $\mathbf{S}_\lambda(\boldsymbol{\theta}^0) = \lim_{n \rightarrow \infty} \mathbf{S}_n(\boldsymbol{\theta}^0)$ .

**Demostración.** Aplicando el desarrollo de Taylor a  $\mathbf{p}(\hat{\boldsymbol{\theta}}_\phi)$  entorno a  $\boldsymbol{\theta}^0$ ,

$$\begin{aligned} \sqrt{n} \left( \mathbf{p}(\hat{\boldsymbol{\theta}}_\phi) - \mathbf{p}(\boldsymbol{\theta}^0) \right) &= \left( \frac{\partial \mathbf{p}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \sqrt{n} (\hat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}^0) + \sqrt{n} \left\| \hat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}^0 \right\| \boldsymbol{\alpha}_2(\hat{\boldsymbol{\theta}}_\phi; \hat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}^0) \\ &= \mathbf{A}(\boldsymbol{\theta}^0) \sqrt{n} (\hat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}^0) + \sqrt{n} \left\| \hat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}^0 \right\| \boldsymbol{\alpha}_2(\hat{\boldsymbol{\theta}}_\phi; \hat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}^0) \end{aligned} \quad (2.24)$$

donde

$$\boldsymbol{\alpha}_2 : \mathbb{R}^p \rightarrow \mathbb{R}^{n_{\text{celdas}} \times n_{\text{celdas}}} \text{ verifica que } \boldsymbol{\alpha}_2(\boldsymbol{\theta}; \boldsymbol{\theta} - \boldsymbol{\theta}^0) \rightarrow \mathbf{0} \text{ cuando } \boldsymbol{\theta} \rightarrow \boldsymbol{\theta}^0 \quad (2.25)$$

y por Teorema 2.3 se tiene el resultado enunciado. ■

## 2.4. Aplicación numérica

Los datos considerados evalúan el nivel de dolor y los medicamentos necesarios de pacientes operados de úlcera de duodeno (Price et al. [93] y Williams y Grizzle [115]). Las respuestas de cada paciente son: para el nivel de dolor ninguno/ leve/ moderado y para la frecuencia de medicación nunca/ rara vez/ ocasionalmente/ regularmente. Dale [36] analizó estos datos presentados en la Tabla 2.1 para un total de 1013 pacientes varones. Varios modelos basados en “cross-ratios” globales, apropiados para esta situación en que las respuestas son cualitativas y ordenadas, fueron ajustados mediante máxima verosimilitud. En este apartado, nuestro objetivo es el de utilizar estos mismos datos para comparar los nuevos procedimientos de estimación basados en  $\phi$ -divergencias introducidos en este capítulo con el de máxima verosimilitud.

Sea  $\mathbf{Z}(\mathbf{x}) = (Z_1(\mathbf{x}), Z_2(\mathbf{x}))$  un vector de variables aleatorias discretas con  $Z_1(\mathbf{x})$  el nivel de dolor (ninguno/ leve y moderado) y  $Z_2(\mathbf{x})$  la frecuencia en la administración de la medicación (nunca/ rara vez/ ocasionalmente/ regularmente), dependiendo de un vector de variables explicativas  $\mathbf{x}$  que en este caso es un vector de unos y ceros, donde los unos los encontramos en la posición  $i$ , con  $i = 1, \dots, 4$  para representar el tipo de operación a la que ha sido sometido un paciente.

Representamos los diferentes tipos de operaciones con:

**VP**=vagotomía y procedimiento de drenaje

**VA**=vagotomía y antrectomía discal

**VH**=vagotomía y hemigastrectomía

**RA**=resección

Tabla 2.1. Número de pacientes, clasificados por el tipo de operación, nivel de dolor y medicación necesaria

Tipo de operación	Nivel de dolor	Medicación necesaria			
		Nunca	Rara vez	Ocasionalmente	Regularmente
<i>VP</i>	Ninguno	170	7	8	0
	Leve	18	5	8	3
	Moderado	7	0	4	14
<i>VA</i>	Ninguno	170	7	5	2
	Leve	22	7	8	1
	Moderado	8	1	8	9
<i>VH</i>	Ninguno	176	8	5	2
	Leve	26	6	5	5
	Moderado	14	3	2	9
<i>RA</i>	Ninguno	181	6	6	2
	Leve	17	12	7	3
	Moderado	10	2	3	11

Se consideran varios modelos CRG (descritos en la Tabla 2.2) de la forma

$$\log(\psi_{ijx}) = \Delta + \alpha_{ia} + \beta_{ja} + \delta_{ij} - \gamma^T \mathbf{x} \quad (2.26)$$

$i = 1, 2$ ,  $j = 1, 2, 3$ , con  $\Delta$  constante,  $\alpha_{1a}$  parámetro de asociación entre dolor ninguno/leve,  $\beta_{1a}$  parámetro de asociación entre medicación nunca/rara vez y  $\beta_{2a}$  parámetro de asociación entre medicación rara vez/ocasionalmente; con restricciones de unicidad simples  $\alpha_{2,a} = \beta_{3,a} = 0$ ;  $\delta_{i,3} = 0$ ,  $i = 1, 2$ ;  $\delta_{2,j} = 0$ ,  $j = 1, 2, 3$ . Entonces  $\gamma = \mathbf{0}$  indica que la asociación no depende del tipo de operación; los  $\delta_{ij}$  cero indican que no hay interacción entre el nivel de dolor y la medicación.

Las probabilidades marginales acumuladas se ajustan utilizando modelos de regresión



logística

$$\begin{aligned}\eta_{i\mathbf{x}} &= \exp(\alpha_{1i} - \mathbf{x}^T \boldsymbol{\beta}_1) / [1 + \exp(\alpha_{1i} - \mathbf{x}^T \boldsymbol{\beta}_1)] & i = 1, 2 \\ \xi_{j\mathbf{x}} &= \exp(\alpha_{2j} - \mathbf{x}^T \boldsymbol{\beta}_2) / [1 + \exp(\alpha_{2j} - \mathbf{x}^T \boldsymbol{\beta}_2)] & j = 1, 2, 3\end{aligned}\quad (2.27)$$

para los niveles de dolor y las necesidades de medicación, respectivamente.

Tabla 2.2. Diferentes modelos de asociación para modelos CRG

Modelos marginales que dependen del tipo de operación	
Modelo 1	$\Delta + \alpha_{ia} + \beta_{ja} + \delta_{ij} - \boldsymbol{\gamma}^T \mathbf{x}$
Modelo 2	$\Delta + \alpha_{ia} + \beta_{ja} - \boldsymbol{\gamma}^T \mathbf{x}$
Modelo 3	$\Delta + \alpha_{ia} + \beta_{ja} + \delta_{ij}$
Modelo 4	$\Delta + \alpha_{ia} - \boldsymbol{\gamma}^T \mathbf{x}$
Modelo 5	$\Delta + \beta_{ja} - \boldsymbol{\gamma}^T \mathbf{x}$
Modelos marginales independientes del tipo de operación	
Modelo 6	$\Delta + \alpha_{ia} + \beta_{ja} + \delta_{ij} - \boldsymbol{\gamma}^T \mathbf{x}$
Modelo 7	$\Delta + \beta_{ja} - \boldsymbol{\gamma}^T \mathbf{x}$
Modelo 8	$\Delta + \beta_{ja}$
Modelo 9	$\Delta + \beta_{ja} - \gamma_1 - \gamma_2$
Modelo 10	$\Delta + \beta_{ja} - \gamma_3$

Los resultados asintóticos que se han obtenido son independientes de la función  $\phi$ . Ahora bien, ese comportamiento equivalente para muestras grandes de los estimadores propuestos no se tiene porque mantener cuando se consideran muestras pequeñas e incluso moderadas. Puesto que el EMV propuesto por Dale [36] constituye un caso particular de los nuevos estimadores de mínima  $\phi$ -divergencia parece necesario el estudiar el comportamiento de estos últimos para muestras pequeñas ya que si se mantuviera la equivalencia para cualquier  $\phi$  carecería de sentido práctico la nueva familia de estimadores introducido. Por otra parte al no ser posible el considerar la familia de familias de divergencia que constituyen las  $\phi$ -divergencia para llevar a cabo el estudio de simulación vamos a elegir la subfamilia paramétrica introducida por Cressie y Read [33] que

se obtiene considerando  $\phi = \phi_{(\lambda)}$  definida por

$$\phi_{(\lambda)}(x) = \frac{1}{\lambda(\lambda+1)} (x^{\lambda+1} - x - \lambda(x-1)); \lambda \neq 0, \lambda \neq -1 \quad (2.28)$$

donde

$$\phi_{(0)}(x) = \lim_{\lambda \rightarrow 0} \phi_{(\lambda)}(x) = x \log x - x + 1,$$

$$\phi_{(-1)}(x) = \lim_{\lambda \rightarrow -1} \phi_{(\lambda)}(x) = -\log x + x + 1,$$

En las Tablas de la 2.3 a la 2.12 aparece la estimación de los parámetros de los diferentes modelos tanto por máxima verosimilitud ( $\lambda = 0$ ) que es el método de estimación que propone Dale [36] como por mínima  $\phi_{(\lambda)}$ -divergencia para valores de  $\lambda = -1/2, 0, 2/3, 1, 2$  y  $3$ .

Junto a los estimadores aparecen los errores estándar ( $E.E.$ ), obtenidos como la raíz cuadrada de los valores de la diagonal de la matriz hessiana obtenida por las subrutinas de SAS NLPFDD y NLPTR [105] utilizadas para optimizar la  $\phi$ -divergencia y obtener los estimadores de mínima  $\phi$ -divergencia. Estas subrutinas no sólo han necesitado la programación de la expresión de la  $\phi$ -divergencia para modelos CRG sino también la de sus primeras derivadas cuyas expresiones se pueden ver en el Anexo. Los valores iniciales de los parámetros para comenzar las subrutinas han sido las mismas que propuso Dale y que han sido descritas en la Sección 2.2.2.

Puesto que la comparación de los errores estándar para los diferentes estimadores resulta difícil debido al gran número de parámetros de los modelos, se ha añadido en la última fila el valor medio de los errores estándar ( $MEE$ ) de los estimadores de todos los parámetros para cada estimador  $\hat{\theta}_{(\lambda)} \equiv \hat{\theta}_{\phi_{(\lambda)}}$  con  $\lambda = -1/2, 0, 2/3, 1, 2$  y  $3$ .

Los mejores estimadores de los parámetros para los modelos cuyos modelos marginales dependen del tipo de operación, **VP**, **VA**, **VH** y **RA** son  $\hat{\theta}_{(2)}$  (3 veces),  $\hat{\theta}_{(1)}$  (1 vez) y  $\hat{\theta}_{(3)}$  (1 vez). En cuanto a los modelos cuyos modelos marginales son independientes del tipo de operación el mejor estimador es  $\hat{\theta}_{(3)}$  y el segundo mejor  $\hat{\theta}_{(2)}$ . Por ello, una buena alternativa para cualquier modelo es  $\hat{\theta}_{(2)}$  ya que excepto para el Modelo 1 que es la tercera mejor opción para el resto de modelos siempre es la mejor alternativa o la segunda mejor opción. Aunque también es una buena elección  $\hat{\theta}_{(1)}$  o  $\hat{\theta}_{(3)}$  ya que para todos los modelos siempre ocupan una de las tres mejores opciones. Lo que es claro es que

el EMV propuesto por Dale ( $\hat{\theta}_{(0)}$ ) nunca aparece ni siquiera como uno de los mejores.

Tabla 2.3. Ajuste del Modelo 1 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = -1/2$		$\lambda = 0$		$\lambda = 2/3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.2014	0.2380	3.2043	0.2144	3.1680	0.2026
$\alpha_{1a}^a$	0.1511	0.4693	0.0037	0.3931	-0.0221	0.3639
$\beta_{1a}^a$	-1.3970	0.3009	-1.2344	0.2927	-1.1173	0.2915
$\beta_{2a}^a$	-0.9123	0.2780	-0.8181	0.2711	-0.7400	0.2729
$\gamma_1$	-0.8119	0.3940	-0.6158	0.3823	-0.5126	0.3788
$\gamma_2$	-0.8338	0.4105	-0.6657	0.4084	-0.5840	0.4064
$\gamma_3$	0.0227	0.3372	0.1694	0.3380	0.2361	0.3396
$\gamma_4$	-0.6624	0.1187	-0.5232	0.1187	-0.4697	0.1197
$\alpha_{11}$	1.3119	0.0748	0.7024	0.0745	1.1134	0.0729
$\alpha_{12}$	2.4012	0.0054	1.7840	0.0118	2.1904	0.0047
$\alpha_{21}$	1.2606	0.0826	1.6187	0.0816	1.4393	0.0799
$\alpha_{22}$	1.7285	0.0677	2.0949	0.0648	1.9222	0.0603
$\alpha_{23}$	2.5769	0.0696	2.9311	0.0672	2.7525	0.0676
$\beta_{11}$	0.1348	0.1695	-0.4425	0.1690	-0.0096	0.1679
$\beta_{12}$	0.2906	0.1742	-0.3110	0.1741	0.1169	0.1755
$\beta_{13}$	0.2733	0.1491	-0.3291	0.1491	0.0914	0.1496
$\beta_{14}$	0.1887	0.0488	-0.4130	0.0495	0.0113	0.0535
$\beta_{21}$	-0.1006	0.1891	0.2890	0.1855	0.1383	0.1823
$\beta_{22}$	-0.2088	0.2013	0.1795	0.1991	0.0328	0.1983
$\beta_{23}$	-0.3391	0.1725	0.0482	0.1710	-0.1055	0.1702
$\beta_{24}$	-0.1774	0.0597	0.2040	0.0624	0.0449	0.0666
$\delta_{11}$	0.0985	0.4560	0.2317	0.3858	0.2546	0.3602
$\delta_{12}$	-0.2102	0.4260	-0.0619	0.3556	-0.0347	0.3307
<i>MEE</i>	0.2127		0.2009		0.1963	

Tabla 2.3 (continuación). Ajuste del Modelo 1 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.1721	0.1993	3.1677	0.1944	3.1040	0.1948
$\alpha_{1a}^a$	-0.0186	0.3573	0.0005	0.3487	0.0270	0.3505
$\beta_{1a}^a$	-1.0724	0.2919	-0.9685	0.2941	-0.8898	0.2966
$\beta_{2a}^a$	-0.7070	0.2752	-0.6244	0.2831	-0.5558	0.2904
$\gamma_1$	-0.4584	0.3780	-0.3535	0.3768	-0.3397	0.3759
$\gamma_2$	-0.5337	0.4068	-0.4301	0.4131	-0.4104	0.4232
$\gamma_3$	0.2799	0.3409	0.3679	0.3474	0.3720	0.3568
$\gamma_4$	-0.4317	0.1211	-0.3567	0.1295	-0.3556	0.1422
$\alpha_{11}$	0.9847	0.0718	1.0320	0.0679	0.8819	0.0636
$\alpha_{12}$	2.0602	0.0129	2.1046	0.0271	1.9533	0.0367
$\alpha_{21}$	1.5323	0.0789	1.5364	0.0754	1.6156	0.0713
$\alpha_{22}$	2.0178	0.0579	2.0279	0.0501	2.1117	0.0410
$\alpha_{23}$	2.8466	0.0685	2.8550	0.0731	2.9392	0.0795
$\beta_{11}$	-0.1299	0.1673	-0.0600	0.1657	-0.1907	0.1645
$\beta_{12}$	-0.0035	0.1769	0.0660	0.1833	-0.0672	0.1911
$\beta_{13}$	-0.0329	0.1502	0.0271	0.1526	-0.1104	0.1561
$\beta_{14}$	-0.1103	0.0563	-0.0408	0.0670	-0.1696	0.0793
$\beta_{21}$	0.2432	0.1811	0.2781	0.1784	0.3850	0.1770
$\beta_{22}$	0.1406	0.1984	0.1832	0.2007	0.2922	0.2046
$\beta_{23}$	-0.0011	0.1703	0.0329	0.1719	0.1361	0.1752
$\beta_{24}$	0.1469	0.0690	0.1743	0.086	0.2727	0.0909
$\delta_{11}$	0.2516	0.3545	0.2368	0.3477	0.2159	0.3496
$\delta_{12}$	-0.0383	0.3251	-0.0615	0.3170	-0.0946	0.3156
<i>MEE</i>	<b>0.1961</b>		0.1975		0.2012	

Tabla 2.4. Ajuste del Modelo 2 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = -1/2$		$\lambda = 0$		$\lambda = 2/3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.2102	0.2328	3.1987	0.2161	3.1784	0.2056
$\alpha_{1a}^a$	0.1939	0.2180	0.1644	0.2128	0.1548	0.2089
$\beta_{1a}^a$	-1.3015	0.2908	-1.1186	0.2824	-0.9815	0.2760
$\beta_{2a}^a$	-0.9015	0.2659	-0.8023	0.2577	-0.7179	0.2526
$\gamma_1$	-0.7684	0.3933	-0.5603	0.3772	-0.4141	0.3678
$\gamma_2$	-0.7863	0.4106	-0.6294	0.4023	-0.5072	0.3925
$\gamma_3$	0.0805	0.3372	0.2123	0.3312	0.3000	0.3242
$\gamma_4$	-0.6657	0.1179	-0.5469	0.1101	-0.4729	0.0942
$\alpha_{11}$	1.4274	0.0749	1.0812	0.0746	1.0833	0.0734
$\alpha_{12}$	2.5104	0.0116	2.1566	0.0156	2.1539	0.0111
$\alpha_{21}$	1.2351	0.0827	1.4330	0.0817	1.4409	0.0802
$\alpha_{22}$	1.6948	0.0683	1.9047	0.0654	1.9223	0.0610
$\alpha_{23}$	2.5323	0.0679	2.7335	0.0663	2.7466	0.0668
$\beta_{11}$	0.2505	0.1695	-0.0621	0.1688	-0.0385	0.1675
$\beta_{12}$	0.4070	0.1739	0.0698	0.1737	0.0903	0.1748
$\beta_{13}$	0.3926	0.1488	0.0509	0.1485	0.0590	0.1482
$\beta_{14}$	0.3057	0.0481	-0.0265	0.0484	-0.0004	0.0510
$\beta_{21}$	-0.1387	0.1893	0.0972	0.1857	0.1379	0.1827
$\beta_{22}$	-0.2417	0.2011	-0.0095	0.1990	0.0376	0.1979
$\beta_{23}$	-0.3671	0.1723	-0.1422	0.1707	-0.1130	0.1689
$\beta_{24}$	-0.2169	0.0592	0.0174	0.0614	0.0616	0.0631
<i>MEE</i>	0.1778		0.1738		0.1699	

Tabla 2.4 (continuación). Ajuste del Modelo 2 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.1594	0.2018	3.1370	0.1935	3.1163	0.1896
$\alpha_{1a}^a$	0.1536	0.2075	0.1551	0.2049	0.1579	0.2042
$\beta_{1a}^a$	-0.9308	0.2734	-0.8140	0.2671	-0.7300	0.2639
$\beta_{2a}^a$	-0.6851	0.2506	-0.6079	0.2454	-0.5539	0.2413
$\gamma_1$	-0.3724	0.3647	-0.2560	0.3580	-0.1804	0.3532
$\gamma_2$	-0.4714	0.3890	-0.3680	0.3847	-0.3030	0.3872
$\gamma_3$	0.3209	0.3214	0.3910	0.3160	0.4379	0.3154
$\gamma_4$	-0.4569	0.0870	-0.3953	0.0734	-0.3515	0.0743
$\alpha_{11}$	0.9217	0.0725	0.7484	0.0692	1.0710	0.0656
$\alpha_{12}$	1.9903	0.0017	1.8111	0.0219	2.1282	0.0326
$\alpha_{21}$	1.5520	0.0793	1.6667	0.0761	1.4788	0.0723
$\alpha_{22}$	2.0372	0.0586	2.1607	0.0505	1.9797	0.0411
$\alpha_{23}$	2.8605	0.0677	2.9833	0.0795	2.8031	0.0795
$\beta_{11}$	-0.1920	0.1668	-0.0053	0.1630	-0.0053	0.1630
$\beta_{12}$	-0.0624	0.1760	0.1271	0.1868	0.1271	0.1868
$\beta_{13}$	-0.0998	0.1481	0.0668	0.1469	0.0668	0.1469
$\beta_{14}$	-0.1489	0.0529	0.0786	0.0672	0.0786	0.0672
$\beta_{21}$	0.2630	0.1815	0.2612	0.1778	0.2612	0.1778
$\beta_{22}$	0.1617	0.1980	0.1537	0.2052	0.1537	0.2052
$\beta_{23}$	0.0071	0.1681	-0.0238	0.1648	-0.0238	0.1648
$\beta_{24}$	0.1891	0.0640	0.1997	0.0753	0.1997	0.0753
MEE	0.1681		<b>0.1665</b>		0.1670	

Tabla 2.5. Ajuste del Modelo 3 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = -1/2$		$\lambda = 0$		$\lambda = 2/3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.7635	0.3303	3.5589	0.3126	3.4161	0.3010
$\alpha_{1a}^a$	0.1399	0.4717	-0.0161	0.3909	-0.0077	0.3666
$\beta_{1a}^a$	-1.4145	0.2983	-1.2095	0.2851	-1.0674	0.2790
$\beta_{2a}^a$	-0.8934	0.2753	-0.7589	0.2608	-0.6507	0.2556
$\alpha_{11}$	0.8366	0.0736	0.6786	0.0739	0.9749	0.0725
$\alpha_{12}$	1.9255	0.0838	1.7606	0.0025	2.0527	0.0101
$\alpha_{21}$	1.4926	0.0838	1.6208	0.0831	1.4952	0.0810
$\alpha_{22}$	1.9638	0.0685	2.0970	0.0665	1.9760	0.0623
$\alpha_{23}$	2.8102	0.0701	2.9295	0.0665	2.8020	0.0660
$\beta_{11}$	-0.3277	0.1678	-0.4582	0.1677	-0.1407	0.1667
$\beta_{12}$	-0.1743	0.1736	-0.3243	0.1734	-0.0107	0.1752
$\beta_{13}$	-0.2374	0.1736	-0.3738	0.1485	-0.0507	0.1490
$\beta_{14}$	-0.2792	0.0500	-0.4274	0.0499	-0.1148	0.0536
$\beta_{21}$	0.1223	0.1902	0.2891	0.1856	0.1976	0.1821
$\beta_{22}$	0.0185	0.2019	0.1778	0.1987	0.0848	0.1974
$\beta_{23}$	-0.1246	0.1731	0.0476	0.1703	-0.0371	0.1693
$\beta_{24}$	0.0527	0.0596	0.2050	0.0623	0.1013	0.0667
$\delta_{11}$	0.1295	0.4599	0.2758	0.3819	0.2758	0.3596
$\delta_{12}$	-0.1975	0.4299	-0.0344	0.3501	-0.0282	0.3265
<i>MEE</i>	0.1969		0.1805		0.1758	

Tabla 2.5 (continuación). Ajuste del Modelo 3 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.3652	0.2970	3.2513	0.2883	3.1730	0.2833
$\alpha_{1a}^a$	0.0096	0.3608	0.0706	0.3513	0.1232	0.3451
$\beta_{1a}^a$	-1.017	0.2775	-0.9073	0.2755	-0.8355	0.2750
$\beta_{2a}^a$	-0.6087	0.2548	-0.5080	0.2550	-0.4346	0.2558
$\alpha_{11}$	0.3992	0.0714	0.7866	0.0670	1.2397	0.0612
$\alpha_{12}$	1.4759	0.0149	1.8620	0.0269	2.3155	0.0371
$\alpha_{21}$	1.8512	0.0797	1.6570	0.0745	1.4099	0.0679
$\alpha_{22}$	2.3341	0.0597	2.1463	0.0502	1.9055	0.0378
$\alpha_{23}$	3.1585	0.0669	2.9684	0.0724	2.7271	0.0795
$\beta_{11}$	-0.7069	0.1661	-0.2939	0.1649	0.1835	0.1641
$\beta_{12}$	-0.5782	0.1768	-0.1696	0.1835	0.3000	0.1909
$\beta_{13}$	-0.6139	0.1495	-0.1920	0.1519	0.2905	0.1551
$\beta_{14}$	-0.6822	0.0564	-0.2707	0.0669	0.2023	0.0786
$\beta_{21}$	0.5677	0.1807	0.4120	0.1775	0.1994	0.1754
$\beta_{22}$	0.4551	0.1976	0.2963	0.1999	0.0762	0.2033
$\beta_{23}$	0.3365	0.1698	0.1875	0.1740	-0.0229	0.1809
$\beta_{24}$	0.4673	0.0700	0.3009	0.0817	0.0772	0.0954
$\delta_{11}$	0.2652	0.3549	0.2268	0.3494	0.1963	0.3478
$\delta_{12}$	-0.0394	0.3210	-0.0859	0.3120	-0.1291	0.3059
<i>MEE</i>	0.1750		<b>0.1749</b>		0.1758	



Tabla 2.6. Ajuste del Modelo 4 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = -1/2$		$\lambda = 0$		$\lambda = 2/3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	2.8527	0.1030	2.8485	0.1102	2.8423	0.1106
$\alpha_{1a}^a$	-0.0148	0.2290	0.0163	0.2160	0.0390	0.2089
$\gamma_1$	-0.5294	0.4125	-0.1839	0.3874	-0.0423	0.3702
$\gamma_2$	-0.1082	0.4291	-0.0563	0.4125	-0.0106	0.4005
$\gamma_3$	0.6693	0.3550	0.6678	0.3420	0.6663	0.3291
$\gamma_4$	-0.1126	0.1145	-0.0897	0.1160	-0.0761	0.1073
$\alpha_{11}$	0.8231	0.0777	0.9905	0.0769	0.9823	0.0755
$\alpha_{12}$	1.9380	0.0054	2.0738	0.0224	2.0458	0.0217
$\alpha_{21}$	1.5578	0.0837	1.4849	0.0812	1.5084	0.0782
$\alpha_{22}$	1.9902	0.0715	1.9347	0.0658	1.9691	0.0600
$\alpha_{23}$	2.8602	0.0727	2.7816	0.0680	2.8049	0.0706
$\beta_{11}$	-0.3688	0.1736	-0.1413	0.1707	-0.1100	0.1674
$\beta_{12}$	-0.2264	0.1777	-0.0449	0.1755	-0.0361	0.1743
$\beta_{13}$	-0.2183	0.1809	-0.0307	0.1499	-0.0184	0.1484
$\beta_{14}$	-0.3221	0.0460	-0.1212	0.0480	-0.0889	0.0509
$\beta_{21}$	0.2167	0.1898	0.2025	0.1829	0.2781	0.1767
$\beta_{22}$	0.1299	0.2008	0.0900	0.1963	0.1441	0.1930
$\beta_{23}$	-0.0186	0.1730	-0.0500	0.1686	0.0102	0.1644
$\beta_{24}$	0.1577	0.0575	0.1165	0.0608	0.1754	0.0632
<i>MEE</i>	0.1644		0.1606		0.1564	

Tabla 2.6 (continuación). Ajuste del Modelo 4 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	2.8457	0.1089	2.8377	0.0992	2.8399	0.0852
$\alpha_{1a}^a$	0.0483	0.2066	0.0686	0.2020	0.0819	0.1989
$\gamma_1$	0.0004	0.3638	0.0621	0.3483	0.1029	0.3362
$\gamma_2$	0.0115	0.3969	0.0441	0.3903	0.0726	0.3868
$\gamma_3$	0.6682	0.3240	0.6669	0.3125	0.6736	0.3046
$\gamma_4$	-0.0636	0.1035	-0.0493	0.0985	-0.0287	0.1022
$\alpha_{11}$	0.9445	0.0746	0.8775	0.0718	1.0453	0.0686
$\alpha_{12}$	2.0016	0.0768	1.9220	0.0159	2.0831	0.0297
$\alpha_{21}$	1.5365	0.0768	1.6026	0.0724	1.4980	0.0678
$\alpha_{22}$	2.0014	0.0572	2.0778	0.0484	1.9817	0.0382
$\alpha_{23}$	2.8352	0.0735	2.9112	0.0841	2.8178	0.0945
$\beta_{11}$	-0.1335	0.1658	-0.1667	0.1608	0.0266	0.1560
$\beta_{12}$	-0.0670	0.1738	-0.1185	0.1728	0.0599	0.1718
$\beta_{13}$	-0.0478	0.1477	-0.0951	0.1454	0.0873	0.1435
$\beta_{14}$	-0.1079	0.0526	-0.1257	0.0581	0.0820	0.0645
$\beta_{21}$	0.3267	0.1742	0.4423	0.1679	0.3760	0.1625
$\beta_{22}$	0.1836	0.1918	0.2743	0.1895	0.1854	0.1882
$\beta_{23}$	0.0524	0.1626	0.1509	0.1583	0.0691	0.1550
$\beta_{24}$	0.2196	0.0642	0.3292	0.0673	0.2611	0.0706
<i>MEE</i>	0.1546		0.1507		<b>0.1488</b>	

Tabla 2.7. Ajuste del Modelo 5 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = -1/2$		$\lambda = 0$		$\lambda = 2/3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.2971	0.2053	3.2614	0.1843	3.2325	0.1717
$\beta_{1a}^a$	-0.1427	0.1721	0.0591	0.1701	0.0856	0.1677
$\beta_{2a}^a$	0.0079	0.0600	0.2179	0.0611	0.2559	0.0616
$\gamma_1$	-0.8155	0.3900	-0.5984	0.3743	-0.4530	0.3645
$\gamma_2$	-0.7906	0.4071	-0.6560	0.3987	-0.5487	0.3881
$\gamma_3$	0.0625	0.3361	0.1840	0.3281	0.2690	0.3191
$\gamma_4$	-0.6949	0.1176	-0.5851	0.1073	-0.5142	0.0877
$\alpha_{11}$	0.9894	0.0757	0.7157	0.0753	0.7680	0.0740
$\alpha_{12}$	2.0762	0.0173	1.7940	0.0211	1.8406	0.0183
$\alpha_{21}$	1.4586	0.0817	1.6317	0.0811	1.6333	0.0797
$\alpha_{22}$	1.9169	0.0673	2.1033	0.0646	2.1156	0.0603
$\alpha_{23}$	2.7456	0.0660	2.9246	0.0653	2.9331	0.0667
$\beta_{11}$	-1.2772	0.2951	-1.1005	0.2858	-0.9685	0.2808
$\beta_{12}$	-0.8916	0.2704	-0.7951	0.2619	-0.7138	0.2583
$\beta_{13}$	-0.1957	0.1696	-0.4363	0.1689	-0.3636	0.1678
$\beta_{14}$	-0.0393	0.1741	-0.3031	0.1740	-0.2327	0.1752
$\beta_{21}$	-0.0561	0.1491	-0.3233	0.1487	-0.2642	0.1482
$\beta_{22}$	-0.1457	0.0488	-0.4000	0.0486	-0.3191	0.0503
$\beta_{23}$	0.0884	0.1883	0.2970	0.1850	0.3288	0.1821
$\beta_{24}$	-0.0165	0.2004	0.1892	0.1982	0.2251	0.1966
<i>MEE</i>	0.1746		0.1701		0.1659	

Tabla 2.7 (continuación). Ajuste del Modelo 5 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.2270	0.1680	3.1704	0.1640	3.1469	0.1669
$\beta_{1a}^a$	0.1519	0.1664	0.0226	0.1629	0.3053	0.1602
$\beta_{2a}^a$	0.3278	0.0619	0.2137	0.0638	0.5079	0.0682
$\gamma_1$	-0.3988	0.3611	-0.3217	0.3537	-0.2551	0.3483
$\gamma_2$	-0.5037	0.3841	-0.4460	0.3766	-0.3933	0.3739
$\gamma_3$	0.3049	0.3154	0.3476	0.3078	0.4026	0.3047
$\gamma_4$	-0.4846	0.0785	-0.4565	0.0590	-0.4159	0.0573
$\alpha_{11}$	0.6774	0.0731	0.9896	0.0699	0.5316	0.0663
$\alpha_{12}$	1.7476	0.0148	2.0528	0.0152	1.5883	0.0274
$\alpha_{21}$	1.6890	0.0789	1.5310	0.0759	1.7878	0.0721
$\alpha_{22}$	2.1754	0.0581	2.0273	0.0505	2.2920	0.0420
$\alpha_{23}$	2.9922	0.0679	2.8437	0.0732	3.1089	0.0798
$\beta_{11}$	-0.9206	0.2791	-0.8085	0.2757	-0.7273	0.2750
$\beta_{12}$	-0.6832	0.2573	-0.6103	0.2555	-0.5602	0.2541
$\beta_{13}$	-0.4470	0.1672	-0.1174	0.1656	-0.5602	0.1642
$\beta_{14}$	-0.3146	0.1763	0.0193	0.1812	-0.4222	0.1868
$\beta_{21}$	-0.3517	0.1479	-0.0309	0.1466	-0.4814	0.1450
$\beta_{22}$	-0.3946	0.0515	-0.0362	0.0557	-0.4511	0.0603
$\beta_{23}$	0.3971	0.1810	0.2732	0.1785	0.5600	0.1767
$\beta_{24}$	0.2955	0.1963	0.1734	0.1971	0.4573	0.1999
<i>MEE</i>	0.1642		<b>0.1614</b>		0.1615	

Tabla 2.8. Ajuste del Modelo 6 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = -1/2$		$\lambda = 0$		$\lambda = 2/3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	1.9862	0.2377	1.2991	0.2198	1.9992	0.2137
$\alpha_{1a}^a$	0.1237	0.4527	-0.0004	0.3918	-0.0363	0.3645
$\beta_{1a}^a$	-13827	0.2950	-1.2319	0.2905	-1.1251	0.2905
$\beta_{2a}^a$	-0.9071	0.2721	-0.8278	0.2671	-0.7644	0.2677
$\gamma_1$	-1.9013	0.3897	-2.4386	0.3793	-1.6264	0.3762
$\gamma_2$	-2.0535	0.4103	-2.5962	0.4073	-1.8020	0.4061
$\gamma_3$	-1.1945	0.3324	-1.7417	0.3359	-0.9345	0.3424
$\gamma_4$	-1.8327	0.1249	-2.1043	0.1272	-1.6337	0.1385
$\alpha_{11}$	1.0961	0.0726	1.0802	0.0722	1.0628	0.0716
$\alpha_{12}$	2.1910	0.1042	2.1644	0.1028	2.1403	0.1014
$\alpha_{21}$	1.4656	0.0806	1.4343	0.0796	1.4047	0.0785
$\alpha_{22}$	1.9342	0.0948	1.9107	0.0939	1.8855	0.0929
$\alpha_{23}$	2.7798	0.1343	2.7486	0.1320	2.7246	0.1300
$\delta_{11}$	0.1391	0.4385	0.2380	0.3836	0.2587	0.3608
$\delta_{12}$	-0.1649	0.4082	-0.0507	0.3530	-0.0280	0.3313
<i>MEE</i>	0.2565		0.2424		0.2377	

Tabla 2.8 (continuación). Ajuste del Modelo 6 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	1.9771	0.2125	3.6089	0.2120	3.4636	0.2130
$\alpha_{1a}^a$	-0.0431	0.3563	-0.0517	0.3396	-0.0547	0.3278
$\beta_{1a}^a$	-1.0848	0.2905	-0.9895	0.2880	-0.9229	0.2816
$\beta_{2a}^a$	-0.7378	0.2682	-0.6697	0.2670	-0.6181	0.2616
$\gamma_1$	-1.6076	0.3763	0.1068	0.3797	0.0097	0.3853
$\gamma_2$	-1.7872	0.4073	-0.0676	0.4149	-0.1483	0.4225
$\gamma_3$	-0.9146	0.3461	0.8226	0.3567	0.7535	0.3642
$\gamma_4$	-0.6335	0.1445	0.0498	0.1603	-0.0593	0.1714
$\alpha_{11}$	1.0550	0.0714	1.0336	0.0704	1.0155	0.0693
$\alpha_{12}$	2.1299	0.1007	2.1021	0.0988	2.0791	0.0971
$\alpha_{21}$	1.3918	0.0780	1.3577	0.0760	1.3293	0.0738
$\alpha_{22}$	1.8739	0.0925	1.8424	0.0907	1.8159	0.0884
$\alpha_{23}$	2.7151	0.1291	2.6920	0.1264	2.6742	0.1239
$\delta_{11}$	0.2597	0.3543	0.2546	0.3417	0.2476	0.3326
$\delta_{12}$	-0.0290	0.3255	-0.0453	0.3154	-0.0677	0.3098
<i>MEF</i>	0.2369		0.2358		<b>0.2348</b>	

Tabla 2.9. Ajuste del Modelo 7 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = -1/2$		$\lambda = 0$		$\lambda = 2/3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	1.4491	0.2069	1.2372	0.1915	1.4146	0.1862
$\beta_{1a}^a$	-1.2663	0.2907	-1.0971	0.2841	-0.9754	0.2804
$\beta_{2a}^a$	-0.8859	0.2658	-0.8024	0.2584	-0.7373	0.2536
$\gamma_1$	-2.5453	0.3851	-2.5443	0.3712	-2.2146	0.3619
$\gamma_2$	-2.6616	0.4060	-2.7056	0.3969	-2.4030	0.3871
$\gamma_3$	-1.8059	0.3306	-1.8491	0.3255	-1.5427	0.3207
$\gamma_4$	-2.5110	0.1229	-2.5869	0.1145	-2.3280	0.1047
$\alpha_{11}$	1.1055	0.0717	1.0853	0.0713	1.0642	0.0708
$\alpha_{12}$	2.1971	0.1029	2.1658	0.1014	2.1367	0.1001
$\alpha_{21}$	1.4728	0.0808	1.4364	0.0796	1.4023	0.0784
$\alpha_{22}$	1.9314	0.0943	1.9079	0.0934	1.8827	0.0925
$\alpha_{23}$	2.7561	0.1314	2.7302	0.1303	2.7089	0.1292
<i>MEE</i>	0.2074		0.2015		0.1971	

Tabla 2.9 (continuación). Ajuste del Modelo 7 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	1.4548	0.1855	2.3408	0.1870	2.1664	0.1913
$\beta_{1a}^a$	-0.9293	0.2787	-0.8189	0.2735	-0.7358	0.2696
$\beta_{2a}^a$	-0.7121	0.2511	-0.6532	0.2420	-0.6172	0.2320
$\gamma_1$	-2.1192	0.3590	-1.1074	0.3549	-1.1980	0.3536
$\gamma_2$	-2.3166	0.3837	-1.3168	0.3788	-1.4105	0.3781
$\gamma_3$	-1.4516	0.3188	-0.4390	0.3147	-0.5147	0.3129
$\gamma_4$	-2.2585	0.1003	-1.3060	0.0904	-1.4256	0.0888
$\alpha_{11}$	1.0546	0.0705	1.0279	0.0694	1.004	0.0681
$\alpha_{12}$	2.1235	0.0995	2.0861	0.0980	2.0516	0.0968
$\alpha_{21}$	1.3874	0.0778	1.3477	0.0757	1.3136	0.0734
$\alpha_{22}$	1.8712	0.0920	1.8405	0.0901	1.8158	0.0876
$\alpha_{23}$	2.7004	0.1286	2.6794	0.1265	2.6630	0.1241
<i>MEE</i>	0.1954		0.1917		<b>0.1897</b>	



Tabla 2.10. Ajuste del Modelo 8 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = -1/2$		$\lambda = 0$		$\lambda = 2/3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.8083	0.2977	3.6119	0.2870	3.4862	0.2787
$\beta_{1a}^a$	-1.2640	0.2864	-1.0623	0.2770	-0.9316	0.2715
$\beta_{2a}^a$	-0.8544	0.2615	-0.7388	0.2474	-0.6586	0.2375
$\alpha_{11}$	1.1053	0.7180	1.0798	0.0713	1.0501	0.0704
$\alpha_{12}$	2.1951	0.1034	2.1599	0.1017	2.1245	0.0997
$\alpha_{21}$	1.4838	0.0810	1.4425	0.0799	1.4425	0.0799
$\alpha_{22}$	1.9434	0.0947	1.9128	0.0935	1.8830	0.0919
$\alpha_{23}$	2.7618	0.1317	2.7281	0.1304	2.7031	0.1288
<i>MEE</i>	0.1660		0.1610		0.1571	

	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.4422	0.2757	3.3415	0.2693	3.2615	0.2650
$\beta_{1a}^a$	-0.8860	0.2701	-0.7838	0.2693	-0.7058	0.2738
$\beta_{2a}^a$	-0.6306	0.2336	-0.5737	0.2236	-0.5434	0.2141
$\alpha_{11}$	1.0375	0.0699	0.9991	0.0681	0.9641	0.0660
$\alpha_{12}$	2.1082	0.0987	2.0622	0.0960	2.0190	0.0937
$\alpha_{21}$	1.3871	0.0776	1.3415	0.0750	1.3010	0.0720
$\alpha_{22}$	1.8703	0.0910	1.8379	0.0881	1.8124	0.0845
$\alpha_{23}$	2.6938	0.1280	2.6722	0.1246	2.6552	0.1201
<i>MEE</i>	0.1556		0.1518		<b>0.1487</b>	

Tabla 2.11. Ajuste del Modelo 9 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = -1/2$		$\lambda = 0$		$\lambda = 2/3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.5969	0.3397	3.4516	0.3236	3.3579	0.3145
$\beta_{1a}^a$	-1.2628	0.2924	-1.0931	0.2843	-0.9710	0.2798
$\beta_{2a}^a$	-0.8820	0.2668	-0.7935	0.2843	-0.7240	0.2509
$\gamma_1$	-0.3928	0.3886	-0.3209	0.3780	-0.2549	0.3730
$\gamma_2$	-0.5095	0.3935	-0.4847	0.3838	-0.4483	0.3710
$\alpha_{11}$	1.1047	0.0718	1.0810	0.0713	1.0543	0.0705
$\alpha_{12}$	2.1931	0.1029	2.1614	0.1016	2.1311	0.1002
$\alpha_{21}$	1.4783	0.0810	1.4404	0.0799	1.4035	0.0786
$\alpha_{22}$	1.9344	0.0945	1.9120	0.0936	1.8881	0.0925
$\alpha_{23}$	2.7571	0.1316	2.7307	0.1304	2.7086	0.1291
<i>MEE</i>	0.2163		0.2104		0.2060	

	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.3241	0.3120	3.2402	0.3098	3.1602	0.3140
$\beta_{1a}^a$	-0.9246	0.2782	-0.8127	0.2754	-0.7213	0.2759
$\beta_{2a}^a$	-0.6974	0.2479	-0.6405	0.2385	-0.6121	0.2296
$\gamma_1$	-0.2288	0.3717	-0.1738	0.3720	-0.1560	0.3796
$\gamma_2$	-0.4320	0.3654	-0.3988	0.3525	-0.3963	0.3448
$\alpha_{11}$	1.0415	0.7003	1.0045	0.0683	0.9706	0.0663
$\alpha_{12}$	2.1169	0.0996	2.0748	0.0978	2.0335	0.0961
$\alpha_{21}$	1.3869	0.0779	1.3410	0.0757	1.2994	0.0733
$\alpha_{22}$	1.8772	0.0919	1.8486	0.0895	1.8262	0.0865
$\alpha_{23}$	2.6996	0.1284	2.6774	0.1262	2.6601	0.1235
<i>MEE</i>	0.2043		0.2006		<b>0.1990</b>	

Tabla 2.12. Ajuste del Modelo 10 (Tabla 2.2) a los datos de la Tabla 2.1

	$\lambda = -1/2$		$\lambda = 0$		$\lambda = 2/3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	4.0070	0.3083	3.8333	0.3048	3.7165	0.2989
$\beta_{1a}^a$	-1.2529	0.2834	-1.0806	0.2793	-0.9600	0.2770
$\beta_{2a}^a$	-0.8716	0.2598	-0.7872	0.2529	-0.7264	0.2483
$\gamma_3$	0.7644	0.3491	0.7617	0.3511	0.7747	0.3566
$\alpha_{11}$	1.1060	0.0717	1.0858	0.0713	1.0648	0.0708
$\alpha_{12}$	2.1990	0.1028	2.1677	0.1014	2.1394	0.1000
$\alpha_{21}$	1.4719	0.0806	1.4350	0.0795	1.3999	0.0782
$\alpha_{22}$	1.9309	0.0940	1.9060	0.0932	1.8800	0.0921
$\alpha_{23}$	2.7546	0.1310	2.7289	0.1303	2.7103	0.1293
<i>MEE</i>	0.1867		0.1849		0.1835	

	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$	
	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>	<i>Estimador</i>	<i>E.E.</i>
$\Delta$	3.6744	0.2958	3.5775	0.2861	3.5053	0.2770
$\beta_{1a}^a$	-0.9148	0.2762	-0.8060	0.2749	-0.7201	0.2743
$\beta_{2a}^a$	-0.7041	0.2462	-0.6567	0.2390	-0.6314	0.2308
$\gamma_3$	0.7848	0.3594	0.8206	0.3664	0.8564	0.3694
$\alpha_{11}$	1.0552	0.0704	1.0290	0.0692	1.0057	0.0677
$\alpha_{12}$	2.1269	0.0993	2.0920	0.0974	2.0592	0.0960
$\alpha_{21}$	1.3843	0.0775	1.3421	0.0752	1.3052	0.0724
$\alpha_{22}$	1.8682	0.0915	1.8371	0.0892	1.8121	0.0864
$\alpha_{23}$	2.7036	0.1286	2.6888	0.1259	2.6784	0.1221
<i>MEE</i>	0.1828		0.1804		<b>0.1774</b>	

## 2.5. Estudio de simulación

Se han ajustado los mismos modelos utilizando la familia de estimadores de mínima  $\phi_{(\lambda)}$ -divergencia para diferentes valores de  $\lambda$  que contiene como caso particular al EMV ( $\lambda = 0$ ) que es el procedimiento de estimación que utilizó Dale [36]. Esto ha permitido comparar por una parte los nuevos estimadores con el EMV y por otra parte encontrar la mejor alternativa, el estimador de mínima  $\phi_{(2)}$ -divergencia para llevar a cabo la estimación de los parámetros en éstos modelos CRG. Puesto que somos conscientes que aunque han sido 10 los modelos ajustados al conjunto de datos al fin y al cabo no deja de ser un ejemplo, parece obligado el llevar a cabo un estudio de simulación para comprobar si efectivamente algún miembro de la familia de estimadores propuesta en este capítulo lleva a un mejor estimador de los parámetros de los modelos CRG.

Se hizo un estudio de simulación ajustando los modelos CRG dados en (2.8) y (2.9) con los modelos de asociación de la Tabla 2.2.

Los datos que se generan consisten en una variable respuesta binaria, discreta y ordenada con cuatro categorías que dependen de un vector de variables explicativas  $\mathbf{x}$  que es un vector de ceros con un uno en la posición  $i = 1, \dots, 4$ . El valor teórico de los parámetros para los diferentes modelos viene dado en la Tabla 2.13. El tamaño de la muestra que se considera es 1000. Se realizan 2000 simulaciones.

Para cada modelo, se ajustan los datos utilizando el estimador de mínima  $\phi$ -divergencia, es decir, se calcula

$$\hat{\boldsymbol{\theta}}_{(\lambda)} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} D_{\phi(\lambda)}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$$

donde  $D_{\phi(\lambda)}$  y  $\phi(\lambda)$  están definidas en (2.17) y (2.28), respectivamente. Se consideran los valores de  $\lambda = -1/2, 0, 2/3, 1, 2$ .

Tabla 2.13. Valores Teóricos de los parámetros

Modelo 1		Modelo 2		Modelo 3		Modelo 4		Modelo 5	
Parámetros de asociación									
$\Delta$	3	$\Delta$	3	$\Delta$	4	$\Delta$	3	$\Delta$	3
$\alpha_{1a}^a$	-0.02	$\alpha_{1a}^a$	0.2	$\alpha_{1a}^a$	0.2	$\alpha_{1a}^a$	-0.1		
$\beta_{1a}^a$	-1	$\beta_{1a}^a$	-1	$\beta_{1a}^a$	-1.5			$\beta_{1a}^a$	-0.5
$\beta_{2a}^a$	-1	$\beta_{2a}^a$	-0.6	$\beta_{2a}^a$	-1			$\beta_{2a}^a$	0
$\gamma_1$	-0.5	$\gamma_1$	-0.3			$\gamma_1$	-0.5	$\gamma_1$	-1
$\gamma_2$	-0.5	$\gamma_2$	-0.4			$\gamma_2$	-0.1	$\gamma_2$	-0.5
$\gamma_3$	0.3	$\gamma_3$	0.5			$\gamma_3$	0.7	$\gamma_3$	0
$\gamma_4$	-0.5	$\gamma_4$	-0.5			$\gamma_4$	-0.1	$\gamma_4$	-1
Parámetros marginales									
$\alpha_{11}$	1	$\alpha_{11}$	1	$\alpha_{11}$	1	$\alpha_{11}$	1	$\alpha_{11}$	1
$\alpha_{12}$	2	$\alpha_{12}$	2	$\alpha_{12}$	2	$\alpha_{12}$	2	$\alpha_{12}$	2
$\alpha_{21}$	1.5	$\alpha_{21}$	1.5	$\alpha_{21}$	1.5	$\alpha_{21}$	1.5	$\alpha_{21}$	1.5
$\alpha_{22}$	2	$\alpha_{22}$	2	$\alpha_{22}$	2	$\alpha_{22}$	2	$\alpha_{22}$	2
$\alpha_{23}$	3	$\alpha_{23}$	3	$\alpha_{23}$	3	$\alpha_{23}$	3	$\alpha_{23}$	3
$\beta_{11}$	-0.2	$\beta_{11}$	-0.01	$\beta_{11}$	-0.5	$\beta_{11}$	-0.3	$\beta_{11}$	-1.5
$\beta_{12}$	-0.01	$\beta_{12}$	0.2	$\beta_{12}$	-0.2	$\beta_{12}$	-0.2	$\beta_{12}$	-1
$\beta_{13}$	-0.03	$\beta_{13}$	0.1	$\beta_{13}$	-0.2	$\beta_{13}$	-0.2	$\beta_{13}$	-0.3
$\beta_{14}$	-0.1	$\beta_{14}$	0.15	$\beta_{14}$	-0.25	$\beta_{14}$	-0.3	$\beta_{14}$	-0.1
$\beta_{21}$	0.3	$\beta_{21}$	0.3	$\beta_{21}$	-0.3	$\beta_{21}$	-0.3	$\beta_{21}$	-0.1
$\beta_{22}$	0.2	$\beta_{22}$	0.2	$\beta_{22}$	0.15	$\beta_{22}$	0.2	$\beta_{22}$	-0.2
$\beta_{23}$	-0.01	$\beta_{23}$	-0.05	$\beta_{23}$	0.02	$\beta_{23}$	-0.1	$\beta_{23}$	-0.1
$\beta_{24}$	0.2	$\beta_{24}$	0.2	$\beta_{24}$	-0.15	$\beta_{24}$	0.2	$\beta_{24}$	-0.05
$\delta_{11}$	0.3			$\delta_{11}$	0.14				
$\delta_{12}$	-0.04			$\delta_{12}$	-0.2				

Tabla 2.13 (continuación). Valores Teóricos de los parámetros

Modelo 6		Modelo 7		Modelo 8		Modelo 9		Modelo 10	
Parámetros de asociación									
$\Delta$	2	$\Delta$	2	$\Delta$	4	$\Delta$	4	$\Delta$	4
$\alpha_{1a}^a$	-0.1								
$\beta_{1a}^a$	-1.4	$\beta_{1a}^a$	-1	$\beta_{1a}^a$	-1	$\beta_{1a}^a$	-1	$\beta_{1a}^a$	-1
$\beta_{2a}^a$	-1	$\beta_{2a}^a$	-1	$\beta_{2a}^a$	-1	$\beta_{2a}^a$	-1	$\beta_{2a}^a$	-0.5
$\gamma_1$	-2	$\gamma_1$	-1			$\gamma_1$	-0.4		
$\gamma_2$	-2	$\gamma_2$	-1			$\gamma_2$	-0.5		
$\gamma_3$	-1	$\gamma_3$	-0.5					$\gamma_3$	1
$\gamma_4$	-1.8	$\gamma_4$	-1						
Parámetros marginales									
$\alpha_{11}$	1	$\alpha_{11}$	1	$\alpha_{11}$	1	$\alpha_{11}$	1	$\alpha_{11}$	1
$\alpha_{12}$	2	$\alpha_{12}$	2	$\alpha_{12}$	2	$\alpha_{12}$	2	$\alpha_{12}$	2
$\alpha_{21}$	1.5	$\alpha_{21}$	1.5	$\alpha_{21}$	1.5	$\alpha_{21}$	1.5	$\alpha_{21}$	1.5
$\alpha_{22}$	2	$\alpha_{22}$	2	$\alpha_{22}$	2	$\alpha_{22}$	2	$\alpha_{22}$	2
$\alpha_{23}$	3	$\alpha_{23}$	3	$\alpha_{23}$	3	$\alpha_{23}$	3	$\alpha_{23}$	3
$\delta_{11}$	0.2								
$\delta_{12}$	-0.1								

A partir de estos valores teóricos y teniendo en cuenta los modelos correspondientes calculamos las probabilidades acumuladas dadas en (2.4), y las probabilidades asociadas (2.11) y generamos 1000 multinomiales utilizando las probabilidades calculadas anteriormente.

Para minimizar  $D_{\phi_{(\lambda)}}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\beta}))$ , utilizamos un método cuasi-Newton de optimización, como son las rutinas de SAS NLPTR y NLPFDD. Para ello necesitamos, las primeras derivadas parciales de la función de logverosimilitud con respecto a los parámetros, dadas en el Anexo y unos valores iniciales, para los algoritmos de optimización que son los propuestos por Dale en la Sección 2.2.2.

La comparación de los diferentes procedimientos se lleva a cabo utilizando las si-

guientes medidas:

El **valor promedio** de los parámetros estimados como

$$\sum_{k=1}^{2000} \frac{\hat{\theta}_{(\lambda),i}^k}{2000}$$

donde  $\hat{\theta}_{(\lambda),i}^k$  es el estimador de mínima  $\phi_{(\lambda)}$ -divergencia del parámetro  $i$  – *ésimo*.

El **error cuadrático medio** ( $ECM$ ),

$$ECM_i = \sqrt{\sum_{k=1}^{2000} \frac{(\hat{\theta}_{(\lambda),i}^k - \theta_i^0)^2}{2000}}$$

donde  $\theta_i^0$  es el verdadero valor del parámetro  $i$  – *ésimo*.

Debido a la gran cantidad de parámetros estimados en cada modelo, también se calcula.

La **media del error cuadrático medio** ( $MECM$ ),

$$MECM = \sum_{i=1}^p \frac{ECM_i}{p}$$

donde  $p$  es el número de parámetros del modelo.

Las Tablas 2.14, 2.16, 2.18, 2.20, 2.22, 2.24, 2.26, 2.28, 2.30, 2.32 presentan los valores promedio de los parámetros estimados para el Modelo 1,..., Modelo 10 respectivamente. Y las Tablas 2.15, 2.17, 2.19, 2.21, 2.23, 2.25, 2.27, 2.28, 2.31, 2.33 presentan los ECM y la MECM de los parámetros  $\alpha_1$ ,  $\beta_1$ ,  $\alpha_2$ ,  $\beta_2$ ,  $\Delta$ ,  $\alpha_a$ ,  $\beta_a$ ,  $\delta$ ,  $\gamma$ .

Tabla 2.14. Valores promedio de los parámetros del Modelo 1

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de Asociación					
$\Delta$	3.0519	2.9809	2.9791	2.9670	2.9524
$\alpha_{1a}^a$	0.1013	0.0909	0.0392	0.0391	0.0467
$\beta_{1a}^a$	-1.1625	-0.9809	-0.9127	-0.8701	-0.8072
$\beta_{2a}^a$	-1.1493	-1.0098	-0.9457	-0.9125	-0.8647
$\gamma_1$	-0.6563	-0.4879	-0.4531	-0.4155	-0.3469
$\gamma_2$	-0.6744	-0.5394	-0.4443	-0.4046	-0.3347
$\gamma_3$	0.1871	0.2966	0.3620	0.3859	0.4451
$\gamma_4$	-0.6587	-0.5213	-0.4370	-0.3985	-0.3268
Parámetros marginales					
$\alpha_{11}$	1.1057	1.0818	1.1184	1.1143	1.1908
$\alpha_{12}$	2.1110	2.0900	2.1252	2.1213	2.1984
$\alpha_{21}$	1.2121	1.2764	1.3048	1.3298	1.3241
$\alpha_{22}$	1.7104	1.7773	1.8057	1.8351	1.8357
$\alpha_{23}$	2.7284	2.7881	2.8057	2.8278	2.8256
$\beta_{11}$	-0.1236	-0.1119	-0.0671	-0.0579	0.0407
$\beta_{12}$	0.0733	0.0794	0.1215	0.1261	0.2239
$\beta_{13}$	0.0543	0.0587	0.1022	0.1062	0.2037
$\beta_{14}$	-0.0086	-0.0185	0.0429	0.0465	0.1473
$\beta_{21}$	-0.0180	0.0757	0.1289	0.1616	0.1901
$\beta_{22}$	-0.1276	-0.0186	0.0270	0.0643	0.0919
$\beta_{23}$	-0.3444	-0.2253	-0.1771	-0.1347	-0.1030
$\beta_{24}$	-0.1179	-0.0302	0.0358	0.0756	0.1002
$\delta_{11}$	0.1782	0.2030	0.2493	0.2502	0.2536
$\delta_{12}$	-0.1697	-0.1250	-0.0688	-0.0615	-0.0487



Tabla 2.15. Error cuadrático Medio. Modelo 1

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	0.2966	0.2671	0.2559	0.2610	0.2678
$\alpha_{1a}^a$	0.5010	0.4095	0.3602	0.3521	0.3456
$\beta_{1a}^a$	0.4446	0.3305	0.3120	0.3345	0.3517
$\beta_{2a}^a$	0.4077	0.2822	0.2603	0.2717	0.2805
$\gamma_1$	0.4551	0.3920	0.3755	0.3821	0.4103
$\gamma_2$	0.4940	0.4021	0.3904	0.3948	0.4203
$\gamma_3$	0.4711	0.4086	0.3943	0.3971	0.4216
$\gamma_4$	0.3218	0.2110	0.2043	0.2218	0.2590
Parámetros de marginales					
$\alpha_{11}$	0.3841	0.3472	0.3278	0.3170	0.3337
$\alpha_{12}$	0.3878	0.3487	0.3395	0.3282	0.3452
$\alpha_{21}$	0.3666	0.3087	0.2803	0.2584	0.2492
$\alpha_{22}$	0.3689	0.3074	0.2794	0.2578	0.2461
$\alpha_{23}$	0.3664	0.3135	0.2964	0.2758	0.2688
$\beta_{11}$	0.4049	0.3674	0.3474	0.3377	0.3782
$\beta_{12}$	0.4054	0.3845	0.3617	0.3512	0.3810
$\beta_{13}$	0.3958	0.3653	0.3559	0.3476	0.3831
$\beta_{14}$	0.3574	0.3171	0.3201	0.3147	0.3548
$\beta_{21}$	0.4088	0.3263	0.2869	0.2604	0.2281
$\beta_{22}$	0.4220	0.3218	0.2853	0.2578	0.2325
$\beta_{23}$	0.4384	0.3236	0.2948	0.2750	0.2471
$\beta_{24}$	0.3892	0.3015	0.2502	0.2292	0.2069
$\delta_{11}$	0.4900	0.3930	0.3548	0.3466	0.3383
$\delta_{12}$	0.4583	0.3416	0.3170	0.3084	0.2983
MECM	0.4103	0.3379	0.3152	<b>0.3079</b>	0.3151

Tabla 2.16. Valores promedio de los parámetros del Modelo 2

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	3.1267	3.0631	3.0237	3.0114	2.9833
$\alpha_{1a}^a$	0.2154	0.2100	0.2121	0.2145	0.2231
$\beta_{1a}^a$	-1.2148	-1.0222	-0.9030	-0.8642	-0.7828
$\beta_{2a}^a$	-0.7589	-0.6174	-0.5341	-0.5079	-0.4544
$\gamma_1$	-0.4018	-0.2752	-0.1866	-0.1547	-0.0849
$\gamma_2$	-0.5381	-0.3806	-0.2760	-0.2389	-0.1596
$\gamma_3$	0.4265	0.5464	0.6220	0.6482	0.7028
$\gamma_4$	-0.6375	-0.4809	-0.3746	-0.3367	-0.2551
Parámetros de marginales					
$\alpha_{11}$	0.7723	0.6902	0.7104	0.7447	0.8637
$\alpha_{12}$	1.7740	1.6943	1.7171	1.7526	1.8750
$\alpha_{21}$	1.2915	1.3778	1.4055	1.4019	1.3596
$\alpha_{22}$	1.7923	1.8804	1.9112	1.9090	1.8705
$\alpha_{23}$	2.8126	2.8829	2.9041	2.8993	2.8567
$\beta_{11}$	-0.2647	-0.3284	-0.2881	-0.2450	-0.1030
$\beta_{12}$	-0.0456	-0.1140	-0.0780	-0.0366	-0.1013
$\beta_{13}$	-0.1463	-0.2131	-0.1757	-0.1339	0.0048
$\beta_{14}$	-0.0890	-0.1559	-0.1185	-0.0765	0.0625
$\beta_{21}$	0.0589	0.1761	0.2346	0.2441	0.2345
$\beta_{22}$	-0.0517	0.0691	0.1303	0.1406	0.1329
$\beta_{23}$	-0.3154	-0.1845	-0.1134	-0.0991	-0.0975
$\beta_{24}$	-0.0462	0.0741	0.1351	0.1454	0.1378

Tabla 2.17. Error cuadrático Medio. Modelo 2

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	0.3009	0.2572	0.2435	0.2422	0.2443
$\alpha_{1a}^a$	0.2072	0.1958	0.1896	0.1884	0.1877
$\beta_{1a}^a$	0.4576	0.3296	0.3171	0.3261	0.3619
$\beta_{2a}^a$	0.4062	0.3038	0.2866	0.2885	0.3026
$\gamma_1$	0.4424	0.3847	0.3842	0.3923	0.4227
$\gamma_2$	0.4656	0.3929	0.3933	0.4029	0.4378
$\gamma_3$	0.4735	0.4134	0.4150	0.4207	0.4437
$\gamma_4$	0.2950	0.2113	0.2335	0.2535	0.3108
Parámetros de marginales					
$\alpha_{11}$	0.4662	0.4647	0.4245	0.3919	0.2755
$\alpha_{12}$	0.4702	0.4663	0.4243	0.3901	0.2758
$\alpha_{21}$	0.2996	0.2282	0.2056	0.2019	0.2142
$\alpha_{22}$	0.3013	0.2303	0.2064	0.2033	0.2115
$\alpha_{23}$	0.3030	0.2502	0.2305	0.2266	0.2372
$\beta_{11}$	0.5010	0.4897	0.4348	0.3921	0.2707
$\beta_{12}$	0.4861	0.4779	0.4244	0.3915	0.2747
$\beta_{13}$	0.4872	0.4764	0.4249	0.3942	0.2918
$\beta_{14}$	0.4554	0.4482	0.4032	0.3630	0.2477
$\beta_{21}$	0.3475	0.2518	0.2128	0.2062	0.1935
$\beta_{22}$	0.3592	0.2598	0.2229	0.2157	0.2049
$\beta_{23}$	0.3811	0.2805	0.2381	0.2298	0.2189
$\beta_{24}$	0.3297	0.2321	0.2018	0.1931	0.1808
MECM	0.3922	0.3922	0.3103	0.3007	<b>0.2766</b>

Tabla 2.18. Valores promedio de los parámetros del Modelo 3

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	4.3737	4.0608	3.8675	3.8065	3.6803
$\alpha_{1a}^a$	0.4343	0.2466	0.1710	0.1558	0.1393
$\beta_{1a}^a$	-1.8585	-1.5737	-1.3936	-1.3427	-1.2435
$\beta_{2a}^a$	-1.2457	-1.0196	-0.9080	-0.8732	-0.8061
Parámetros de marginales					
$\alpha_{11}$	0.8802	1.2220	1.3959	1.4153	1.4503
$\alpha_{12}$	1.8867	2.2283	2.400	2.4187	2.4523
$\alpha_{21}$	1.7324	1.5910	1.5248	1.5259	1.5335
$\alpha_{22}$	2.2275	2.0887	2.0276	2.0299	2.0407
$\alpha_{23}$	3.2484	3.0921	3.0220	3.0222	3.0308
$\beta_{11}$	-0.6615	-0.2889	-0.0822	-0.0494	0.0184
$\beta_{12}$	-0.3472	0.0156	0.2125	0.2418	0.3021
$\beta_{13}$	-0.3455	0.0174	0.2142	0.2435	0.3038
$\beta_{14}$	-0.3914	-0.0273	0.1707	0.2006	0.2621
$\beta_{21}$	-0.1391	-0.2261	-0.2401	-0.2189	-0.1632
$\beta_{22}$	0.3343	0.2261	0.1940	0.2086	0.2495
$\beta_{23}$	0.2000	0.0976	0.0701	0.0866	0.1318
$\beta_{24}$	0.0343	-0.0604	-0.0819	-0.0633	-0.0135
$\delta_{11}$	-0.0856	0.1129	0.1879	0.2061	0.2315
$\delta_{12}$	-0.4704	-0.2620	-0.1581	-0.1353	-0.1014

Tabla 2.19. Error cuadrático Medio. Modelo 3

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	0.6190	0.4040	0.3867	0.4065	0.4739
$\alpha_{1a}^a$	0.6720	0.5085	0.4450	0.4316	0.4081
$\beta_{1a}^a$	0.6540	0.5274	0.4397	0.4418	0.4640
$\beta_{2a}^a$	0.5635	0.4157	0.3684	0.3692	0.3823
Parámetros de marginales					
$\alpha_{11}$	1.0745	0.8597	0.7059	0.6774	0.6189
$\alpha_{12}$	1.0843	0.8674	0.7069	0.6785	0.6185
$\alpha_{21}$	0.6334	0.4960	0.3574	0.3297	0.2667
$\alpha_{22}$	0.6264	0.4896	0.3506	0.3243	0.2652
$\alpha_{23}$	0.6442	0.4923	0.3629	0.3404	0.2816
$\beta_{11}$	1.0994	0.8843	0.7344	0.7107	0.6709
$\beta_{12}$	1.0866	0.8620	0.7216	0.7018	0.6612
$\beta_{13}$	1.0735	0.8828	0.7626	0.7472	0.7199
$\beta_{14}$	1.0570	0.8742	0.7497	0.7287	0.7000
$\beta_{21}$	0.6689	0.6163	0.537	0.5029	0.4272
$\beta_{22}$	0.6644	0.5289	0.4078	0.3889	0.3681
$\beta_{23}$	0.7001	0.5486	0.4218	0.4075	0.3881
$\beta_{24}$	0.6124	0.5102	0.3840	0.3520	0.2789
$\delta_{11}$	0.6752	0.5484	0.4695	0.4543	0.4251
$\delta_{12}$	0.6691	0.5202	0.4414	0.4263	0.3982
MECM	0.7830	0.6230	0.5133	0.4958	<b>0.4640</b>

Tabla 2.20. Valores promedio de los parámetros del Modelo 4

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	3.0632	3.0448	3.0340	3.0305	3.0282
$\alpha_{1a}^a$	-0.1270	-0.0933	-0.0647	-0.0530	-0.0244
$\gamma_1$	-0.6489	-0.5181	-0.4280	-0.3952	-0.3162
$\gamma_2$	-0.1285	-0.0764	-0.0234	-0.0014	0.0584
$\gamma_3$	0.7181	0.7350	0.7550	0.7648	0.7957
$\gamma_4$	-0.1319	-0.0751	-0.0240	-0.0018	0.0574
Parámetros de marginales					
$\alpha_{11}$	1.3576	1.3718	1.3475	1.3305	1.2869
$\alpha_{12}$	2.3669	2.3788	2.3519	2.3336	2.2869
$\alpha_{21}$	1.3289	1.3467	1.3855	1.4064	1.4579
$\alpha_{22}$	1.8248	1.8475	1.8914	1.9145	1.9723
$\alpha_{23}$	2.8566	2.8593	2.8912	2.9108	2.9619
$\beta_{11}$	0.0219	0.0636	0.0673	0.0623	0.0496
$\beta_{12}$	0.1290	0.1664	0.1672	0.1612	0.1461
$\beta_{13}$	0.1282	0.1669	0.1679	0.1618	0.1461
$\beta_{14}$	0.0322	0.0727	0.0759	0.0709	0.0581
$\beta_{21}$	-0.5410	-0.4658	-0.3749	-0.3327	-0.2294
$\beta_{22}$	-0.0110	0.0360	0.1067	0.1416	0.2291
$\beta_{23}$	-0.3272	-0.2637	-0.1791	-0.1388	-0.0394
$\beta_{24}$	0.0016	0.0479	0.1180	0.1528	0.2400

Tabla 2.21. Error cuadrático Medio. Modelo 4

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	0.2621	0.2423	0.2323	0.2315	0.2343
$\alpha_{1a}^a$	0.2457	0.2179	0.2111	0.2114	0.2176
$\gamma_1$	0.5129	0.4254	0.4054	0.4082	0.4351
$\gamma_2$	0.4057	0.3751	0.3690	0.3715	0.3961
$\gamma_3$	0.4264	0.3996	0.3913	0.3930	0.4072
$\gamma_4$	0.1784	0.1599	0.1713	0.1814	0.2209
Parámetros de marginales					
$\alpha_{11}$	0.4628	0.4910	0.5059	0.4994	0.4372
$\alpha_{12}$	0.4747	0.5023	0.5122	0.5081	0.4412
$\alpha_{21}$	0.2476	0.2535	0.2556	0.2483	0.2164
$\alpha_{22}$	0.2541	0.2586	0.2573	0.2482	0.2186
$\alpha_{23}$	0.2660	0.2795	0.2814	0.2717	0.2450
$\beta_{11}$	0.4539	0.5021	0.5292	0.5301	0.4919
$\beta_{12}$	0.4617	0.5053	0.5359	0.5364	0.4949
$\beta_{13}$	0.4601	0.5061	0.5387	0.5392	0.5019
$\beta_{14}$	0.4338	0.4854	0.5179	0.5155	0.4777
$\beta_{21}$	0.3546	0.3117	0.2864	0.2751	0.2599
$\beta_{22}$	0.2905	0.2762	0.2720	0.2586	0.2399
$\beta_{23}$	0.3376	0.3017	0.2838	0.2789	0.2730
$\beta_{24}$	0.2707	0.2452	0.2283	0.2247	0.2060
MECM	0.3579	0.3547	0.3571	0.3543	<b>0.3376</b>

Tabla 2.22. Valores promedio de los parámetros del Modelo 5

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	3.9446	3.8081	3.7536	3.7455	3.7264
$\beta_{1a}^a$	-0.8042	-0.5466	-0.4076	-0.3700	-0.2933
$\beta_{2a}^a$	-0.0421	0.0135	0.0531	0.0678	0.0961
$\gamma_1$	-0.4562	-0.3067	-0.1745	-0.1340	-0.0376
$\gamma_2$	0.1779	0.2553	0.3396	0.3639	0.4323
$\gamma_3$	0.6591	0.7571	0.8427	0.8633	0.9278
$\gamma_4$	-0.5068	-0.2725	-0.1221	-0.0788	0.0225
Parámetros de marginales					
$\alpha_{11}$	1.1202	1.2883	1.4183	1.5125	1.6643
$\alpha_{12}$	2.1155	2.2912	2.4210	2.5146	2.6655
$\alpha_{21}$	1.6249	1.5839	1.5752	1.5282	1.4517
$\alpha_{22}$	2.1202	2.0825	2.0776	2.0306	1.9553
$\alpha_{23}$	3.1364	3.0867	3.0701	3.0192	2.9347
$\beta_{11}$	-1.4876	-1.2363	-1.0294	-0.9071	-0.6930
$\beta_{12}$	-0.9425	-0.7341	-0.5221	-0.4396	-0.2399
$\beta_{13}$	-0.2097	-0.0216	0.1342	0.2388	0.4186
$\beta_{14}$	0.0034	0.1918	0.3420	0.4454	0.6192
$\beta_{21}$	-0.0364	-0.0303	0.0032	-0.0263	-0.0642
$\beta_{22}$	-0.1509	-0.1347	-0.0930	-0.1214	-0.1525
$\beta_{23}$	-0.0383	-0.0280	0.0118	-0.0179	-0.0489
$\beta_{24}$	-0.0189	0.0289	0.0620	0.0299	-0.0082



Tabla 2.23. Error cuadrático Medio. Modelo 5

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	1.0365	0.8654	0.8134	0.7962	0.7799
$\beta_{1a}^a$	0.6840	0.4334	0.3599	0.3670	0.3920
$\beta_{2a}^a$	0.4530	0.3428	0.3150	0.3128	0.3130
$\gamma_1$	0.9236	0.8911	0.9655	0.9970	1.0787
$\gamma_2$	0.8735	0.8887	0.9493	0.9701	1.0346
$\gamma_3$	0.8456	0.8796	0.9445	0.9623	1.0224
$\gamma_4$	0.6399	0.7783	0.9123	0.9524	1.0483
Parámetros de marginales					
$\alpha_{11}$	0.6015	0.5983	0.6109	0.6415	0.7203
$\alpha_{12}$	0.6079	0.6092	0.6201	0.6480	0.7230
$\alpha_{21}$	0.4668	0.4308	0.3709	0.3256	0.2420
$\alpha_{22}$	0.4669	0.4288	0.3699	0.3279	0.2450
$\alpha_{23}$	0.4883	0.4442	0.3827	0.3427	0.2722
$\beta_{11}$	0.6985	0.6623	0.6910	0.7318	0.8634
$\beta_{12}$	0.6238	0.6097	0.6512	0.6957	0.5213
$\beta_{13}$	0.5944	0.5913	0.6206	0.6713	0.7890
$\beta_{14}$	0.5827	0.5859	0.6161	0.6634	0.7673
$\beta_{21}$	0.4331	0.3928	0.3456	0.3177	0.2705
$\beta_{22}$	0.4726	0.4358	0.3946	0.3643	0.2887
$\beta_{23}$	0.5016	0.4761	0.4176	0.3629	0.2635
$\beta_{24}$	0.4808	0.4546	0.4030	0.3451	0.2339
MECM	0.6236	0.5900	<b>0.5877</b>	0.5898	0.6085

Tabla 2.24. Valores promedio de los parámetros del Modelo 6

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	1.2687	1.2472	1.9449	2.2675	2.9420
$\alpha_{1a}^a$	0.1406	-0.0432	-0.0261	-0.0403	-0.0383
$\beta_{1a}^a$	-1.7561	-1.7465	-1.2959	-1.2715	-1.1698
$\beta_{2a}^a$	-1.2956	-0.8452	-0.9319	-0.8370	-0.8302
$\gamma_1$	-3.1390	-2.8552	-1.9340	-1.5153	-0.7653
$\gamma_2$	-3.1542	-2.8820	-1.9435	-1.5390	-0.7731
$\gamma_3$	-2.0569	-1.8810	-0.9450	-0.5871	0.1815
$\gamma_4$	-2.9369	-2.6809	-1.7440	-1.3536	-0.5826
Parámetros de marginales					
$\alpha_{11}$	1.0206	1.0104	0.9786	0.9725	0.9447
$\alpha_{12}$	2.0213	2.0324	1.9841	1.9831	1.9533
$\alpha_{21}$	1.5429	1.5276	1.4613	1.4512	1.4048
$\alpha_{22}$	2.0389	1.9827	1.9631	1.9440	1.9109
$\alpha_{23}$	3.0627	2.9827	2.9576	2.9362	2.9010
$\delta_{11}$	-0.0406	0.3183	0.1375	0.1755	0.1631
$\delta_{12}$	-0.3551	-0.3803	-0.1374	-0.1561	-0.1064

Tabla 2.25. Error cuadrático Medio. Modelo 6

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	1.0347	1.3820	1.3630	1.6020	1.8674
$\alpha_{1a}^a$	0.9051	0.6676	0.5623	0.5380	0.4933
$\beta_{1a}^a$	0.6764	1.3091	0.3718	0.5719	0.4536
$\beta_{2a}^a$	0.6023	0.8626	0.3372	0.4677	0.3765
$\gamma_1$	1.6059	1.7566	1.5448	1.8357	2.1041
$\gamma_2$	1.5927	1.7485	1.5426	1.8299	2.1067
$\gamma_3$	1.0687	1.5125	1.8588	2.2732	2.8093
$\gamma_4$	1.5345	1.6251	1.3391	1.6639	1.9932
Parámetros de marginales					
$\alpha_{11}$	0.0767	0.1068	0.0733	0.0792	0.0891
$\alpha_{12}$	0.1032	0.2029	0.0953	0.1113	0.1067
$\alpha_{21}$	0.0971	0.1712	0.0880	0.1050	0.1257
$\alpha_{22}$	0.1098	0.1570	0.1009	0.1102	0.1290
$\alpha_{23}$	0.1752	0.2350	0.1486	0.1565	0.1723
$\delta_{11}$	0.8744	1.0792	0.5352	0.5900	0.4911
$\delta_{12}$	0.8390	0.9695	0.5029	0.5312	0.4533
MECM	0.7531	0.9190	<b>0.6976</b>	0.8310	0.9181

Tabla 2.26. Valores promedio de los parámetros del Modelo 7

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	1.4612	1.7686	2.0322	2.1229	2.2356
$\beta_{1a}^a$	-1.2108	-1.0284	-0.9243	-0.8890	-0.8129
$\beta_{2a}^a$	-1.1861	-1.0250	-0.9378	-0.9089	-0.8478
$\gamma_1$	-1.7684	-1.2690	-0.8956	-0.7664	-0.5710
$\gamma_2$	-1.7679	-1.2680	-0.8945	-0.7662	-0.5711
$\gamma_3$	-1.2220	-0.7529	-0.3988	-0.2774	-0.0967
$\gamma_4$	-1.7675	-1.2672	-0.8921	-0.7630	-0.5655
Parámetros de marginales					
$\alpha_{11}$	1.0217	1.0006	0.9773	0.9669	0.9399
$\alpha_{12}$	2.0218	2.0055	1.9856	1.9768	1.9537
$\alpha_{21}$	1.5468	1.5021	1.4574	1.4384	1.3904
$\alpha_{22}$	2.0434	1.9999	1.9583	1.9408	1.8970
$\alpha_{23}$	3.0713	3.0043	2.9496	2.9282	2.8772

Tabla 2.27. Error cuadrático Medio. Modelo 7

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	1.3441	1.3415	1.3396	1.3741	1.4924
$\beta_{1a}^a$	0.4653	0.2967	0.2794	0.2870	0.3225
$\beta_{2a}^a$	0.4402	0.2793	0.2589	0.2629	0.2869
$\gamma_1$	1.6905	1.5255	1.4882	1.5164	1.6414
$\gamma_2$	1.6681	1.5155	1.4643	1.4971	1.6261
$\gamma_3$	1.5874	1.4465	1.4288	1.4725	1.6154
$\gamma_4$	1.4639	1.2638	1.2537	1.3054	1.4738
Parámetros de marginales					
$\alpha_{11}$	0.0758	0.0705	0.0729	0.0765	0.0915
$\alpha_{12}$	0.1031	0.0963	0.0946	0.0958	0.1035
$\alpha_{21}$	0.0984	0.0810	0.0892	0.0994	0.1345
$\alpha_{22}$	0.1133	0.0974	0.1028	0.1105	0.1390
$\alpha_{23}$	0.1812	0.1481	0.1497	0.1570	0.1858
MECM	0.7693	0.6802	<b>0.6685</b>	0.6879	0.7594

Tabla 2.28. Valores promedio de los parámetros del Modelo 8

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	4.5567	4.0831	3.8175	3.7417	3.5923
$\beta_{1a}^a$	-1.5171	-1.0768	-0.8615	-0.8058	-0.7054
$\beta_{2a}^a$	-1.4860	-1.0796	-0.8894	-0.8424	-0.7614
Parámetros de marginales					
$\alpha_{11}$	1.0147	1.0000	0.9748	0.9635	0.9340
$\alpha_{12}$	2.0104	2.0048	1.9802	1.9692	1.9909
$\alpha_{21}$	1.5366	1.5007	1.4537	1.4341	1.3855
$\alpha_{22}$	2.0347	2.0021	1.9575	1.9395	1.8953
$\alpha_{23}$	3.0532	3.0051	2.9508	2.9309	2.8856

Tabla 2.29. Error cuadrático Medio. Modelo 8

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	1.0009	0.7335	0.5218	0.5554	0.6413
$\beta_{1a}^a$	0.8969	0.7259	0.5013	0.5220	0.5698
$\beta_{2a}^a$	0.8634	0.7094	0.4728	0.4885	0.5224
Parámetros de marginales					
$\alpha_{11}$	0.1145	0.0699	0.0730	0.0773	0.0947
$\alpha_{12}$	0.2014	0.0954	0.0944	0.0967	0.1088
$\alpha_{21}$	0.1629	0.0815	0.0915	0.1024	0.1389
$\alpha_{22}$	0.2075	0.0969	0.1027	0.1109	0.1401
$\alpha_{23}$	0.3147	0.1490	0.1511	0.1580	0.1829
MECM	0.4703	0.3327	<b>0.2511</b>	0.2639	0.2999

Tabla 2.30. Valores promedio de los parámetros del Modelo 9

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	4.6902	4.1089	3.8262	3.7452	3.5927
$\beta_{1a}^a$	-1.6337	-1.1065	-0.8706	-0.8080	-0.6995
$\beta_{2a}^a$	-1.5964	-1.1062	-0.8982	-0.8456	-0.7531
$\gamma_1$	-0.4424	-0.4240	-0.4093	-0.4043	-0.3939
$\gamma_2$	-0.5654	-0.5369	-0.5144	-0.5067	-0.4914
Parámetros de marginales					
$\alpha_{11}$	1.0221	1.0003	0.9760	0.9652	0.9370
$\alpha_{12}$	2.0251	2.0054	1.9824	1.9722	1.9458
$\alpha_{21}$	1.5457	1.5007	1.4563	1.4378	1.3916
$\alpha_{22}$	2.0463	2.0020	1.9605	1.9436	1.9021
$\alpha_{23}$	3.0706	3.0052	2.9541	2.9354	2.8930

Tabla 2.31. Error cuadrático Medio. Modelo 9

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	1.1149	0.7789	0.6333	0.6417	0.7389
$\beta_{1a}^a$	1.0655	0.7665	0.6088	0.6052	0.6721
$\beta_{2a}^a$	1.0287	0.7491	0.5831	0.6736	0.6274
$\gamma_1$	0.4694	0.4286	0.4771	0.4018	0.3955
$\gamma_2$	0.4817	0.4367	0.4156	0.4110	0.4068
Parámetros de marginales					
$\alpha_{11}$	0.0752	0.0699	0.0726	0.0765	0.0927
$\alpha_{12}$	0.1024	0.0952	0.0938	0.0956	0.1059
$\alpha_{21}$	0.0969	0.0810	0.0900	0.1000	0.1339
$\alpha_{22}$	0.1127	0.0972	0.1021	0.1093	0.1356
$\alpha_{23}$	0.1740	0.1484	0.1498	0.1557	0.1778
MECM	0.4721	0.3651	<b>0.3156</b>	0.3170	0.3487

Tabla 2.32. Valores promedio de los parámetros del Modelo 10

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	4.4801	4.0956	3.8351	3.7566	3.6227
$\beta_{1a}^a$	-1.4048	-1.1043	-0.8749	-0.8233	-0.7330
$\beta_{2a}^a$	-0.8123	-0.5115	-0.4181	-0.3636	-0.3225
$\gamma_3$	1.0991	1.0214	0.9701	0.9377	0.0993
Parámetros de marginales					
$\alpha_{11}$	1.0234	1.0056	0.9758	0.9667	0.9373
$\alpha_{12}$	2.0257	2.0167	1.9821	1.9769	1.9477
$\alpha_{21}$	1.5501	1.5133	1.4555	1.4421	1.3910
$\alpha_{22}$	2.0506	1.9992	1.9587	1.9373	1.8970
$\alpha_{23}$	3.0733	2.9986	2.9510	2.9273	2.8845

Tabla 2.33. Error cuadrático Medio. Modelo 10

	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
Parámetros de asociación					
$\Delta$	0.7466	0.4500	0.4032	0.4390	0.5209
$\beta_{1a}^a$	0.6850	0.4700	0.3701	0.3897	0.4323
$\beta_{2a}^a$	0.6010	0.3721	0.3269	0.3466	0.3517
$\gamma_3$	0.4271	0.3682	0.3511	0.3578	0.3573
Parámetros de marginales					
$\alpha_{11}$	0.0756	0.0722	0.0723	0.0762	0.0925
$\alpha_{12}$	0.1035	0.1032	0.0941	0.0970	0.1061
$\alpha_{21}$	0.1003	0.0930	0.0901	0.1000	0.1352
$\alpha_{22}$	0.1154	0.0964	0.1018	0.1125	0.1386
$\alpha_{23}$	0.1787	0.1492	0.1505	0.1592	0.1826
MECM	0.3370	0.2416	<b>0.2178</b>	0.2309	0.2575



El mejor estimador de los parámetros del modelo 1 es  $\hat{\theta}_{(1)}$ . Para los modelos 2, 3 y 4 es  $\hat{\theta}_{(2)}$  seguido de  $\hat{\theta}_{(1)}$ . Sin embargo, para el resto de los modelos el mejor estimador es  $\hat{\theta}_{(\frac{2}{3})}$  seguido de  $\hat{\theta}_{(1)}$  y  $\hat{\theta}_{(2)}$ . Es decir, que en el estudio de simulación nos aparece una nueva alternativa para la estimación de los parámetros de los modelos CRG que es el estimador de mínima  $\phi_{(\frac{2}{3})}$ -divergencia. Un buen compromiso para todas las alternativas es el estimador  $\hat{\theta}_{(1)}$  aunque  $\hat{\theta}_{(\frac{2}{3})}$  y  $\hat{\theta}_{(2)}$  también son buenas alternativas ya que ocupan siempre una de las tres primeras posiciones.

No es la primera vez que el estimador de mínima  $\chi^2$  de Pearson ( $\lambda = 1$ ) aparece como una buena alternativa al estimador de máxima verosimilitud ( $\lambda = 0$ ). En ensayos clínicos, Berkson [17] encontró que es más pequeño el error cuadrático medio asociado del estimador de mínima  $\chi^2$  de Pearson que al EMV. También Pardo et al. [80] llegó a que este estimador era el miembro de la familia de estimadores de mínima divergencia de Cressie y Read más robusto y eficiente para estimar modelos loglineales.

## Capítulo 3

# Extensión de las Ecuaciones de Estimación Generalizadas

A mediados de los 80, se consiguieron importantes avances en la metodología para analizar datos longitudinales discretos cuando Liang y Zeger [61] y Zeger y Liang [116] propusieron la metodología de ecuaciones de estimación generalizadas (EEG). Puesto que los modelos marginales parametrizan de forma separada el modelo para la media de la variable respuesta y el modelo para la asociación de las variables medidas de cada individuo, Liang y Zeger [61] vieron que no era necesaria la distribución conjunta de las observaciones de un individuo para estimar los parámetros de regresión del primer modelo. Esto es una ventaja porque como se ha visto en el capítulo anterior no hay una especificación sencilla de la distribución conjunta de la variable respuesta para modelos marginales cuando la respuesta es discreta. En el método basado en las ecuaciones de estimación generalizadas (EEG) sólo es necesario especificar la verosimilitud para la distribución marginal y la matriz de correlación del vector de medidas repetidas de cada individuo. El método EEG es una extensión del método de cuasi-verosimilitud Wedderburn [112] para modelos lineales generalizados multinomiales, donde se incluye un conjunto de parámetros de ruido para las asociaciones dentro de cada individuo. Extensiones de las EEG que permiten modelizar medidas de asociación tales como la correlación o el cociente de disparidad han sido propuestos por Prentice [92], Lipstiz, Laird

y Harrington [64], Liang, Zeger y Qaqish [62] y Carey, Zeger y Diggle [37].

El objetivo de este capítulo es generalizar la metodología de las EEG proponiendo nuevas formas de estimación de los parámetros de correlación. Puesto que la metodología EEG es una extensión de los modelos lineales generalizados, en la Sección 3.1 se definen estos modelos. A continuación en la Sección 3.2, se introduce la metodología EEG. En la Sección 3.3 se propone una generalización de la metodología EEG y se ilustra con algunos ejemplos en la Sección 3.4. Finalmente, en la Sección 3.5 se lleva a cabo un estudio de simulación para comparar los diferentes procedimientos de estimación propuestos en el capítulo con el procedimiento de las EEG.

### 3.1. Modelo Lineal Generalizado

Esta sección esta dedicada a la definición del Modelo Lineal Generalizado (MLG) que es la base del método de las EEG. Los MLG fueron introducidos por Nelder y Wedderburn [73] como una forma de unificar varios modelos estadísticos tales como regresión lineal, regresión logística y regresión de Poisson entre otros, bajo un único marco teórico.

Sea  $\mathbf{Y}_i = (y_{i1}, \dots, y_{it_i})^T$  el vector respuesta  $t_i \times 1$  y sea  $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{it_i})^T$  la matriz de variables explicativas  $t_i \times p$  del  $i$ -ésimo individuo ( $i = 1, \dots, n$ ), donde  $y_{ij}$  denota la respuesta del sujeto  $i$  en el tiempo o condición  $j$ ,  $j = 1, \dots, t_i$ ,  $i = 1, \dots, n$ ;  $t_i$  es el número de medidas del individuo  $i$  y  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$  denota un vector de variables explicativas asociado con  $y_{ij}$ . Se supone que la densidad marginal de  $y_{ij}$  es

$$f(y_{ij}) = \exp \left[ \frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\gamma)} + c(y_{ij}, \gamma) \right].$$

La media y la varianza de  $y_{ij}$  vienen dadas por

$$E(y_{ij}) = \mu_{ij} = b'(\theta_{ij}), \text{var}(y_{ij}) = v(\mu_{ij})a(\gamma)$$

donde  $v(\mu_{ij}) = b''(\theta_{ij})$  es la función de varianza y  $\gamma$  es un posible parámetro de dispersión.

El Modelo Lineal Generalizado (MLG) relaciona la media de la variable respuesta con un vector de variables explicativas a través de una función enlace

$$\boldsymbol{\eta}_i = \mathbf{X}_i^T \boldsymbol{\beta},$$

donde  $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{it_i})^T$  con  $\eta_{ij} = g(\mu_{ij})$  y  $\eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}$ .  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  es el vector  $p \times 1$  de parámetros desconocidos que deben ser estimados.

En la Tabla 3.1 se muestra las funciones  $a$ ,  $b$  y  $c$  que se deben considerar cuando las variables respuesta son binomiales, Poisson o Normales así como las características de dichas funciones.

Tabla 3.1. Características de algunas distribuciones de la familia exponencial

	Normal	Poisson	Binomial
<i>Rango de <math>y_{ij}</math></i>	$(-\infty, \infty)$	$0(1)\infty$	$0(1)n$
$a()$	$\gamma$	1	1
$b()$	$\frac{1}{2}\theta_{ij}^2$	$e^{\theta_{ij}}$	$\ln(1 + e^{\theta_{ij}})$
$c()$	$-\frac{1}{2}\left(\frac{y_{ij}^2}{\gamma} + \ln(2\pi\gamma)\right)$	$-\ln y_{ij}!$	$\ln\left[\binom{n}{y_{ij}}\right]$
$\mu = E(Y)$	$\theta_{ij}$	$e^{\theta_{ij}}$	$ne^{\theta_{ij}} / (1 + e^{\theta_{ij}})$
<i>Función Varianza</i>	1	$\mu_{ij}$	$\frac{\mu_{ij}(1 - \mu_{ij})}{n}$

Bajo la suposición de independencia, obtenemos el estimador de máxima verosimilitud  $\hat{\boldsymbol{\beta}}$ , del vector  $p \times 1$  de parámetros  $\boldsymbol{\beta}$ , resolviendo las ecuaciones

$$\sum_{i=1}^n \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\beta}} (\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) = \mathbf{0},$$

donde  $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}), \dots, \mu_{it_i}(\boldsymbol{\beta}))^T$ .

Estas ecuaciones son generalmente no lineales y por tanto no tienen solución explícita. Así pues, es necesario un algoritmo de optimización para su resolución. Los métodos de

Newton-Raphson y de puntuaciones de Fisher son los que usualmente se utilizan. Por ejemplo, en el procedimiento de SAS PROC GENMOD el algoritmo de ajuste empieza con unos pocos pasos de Fisher y entonces cambia al método de Newton-Raphson.

Sea  $\mathbf{A}_i$  una matriz diagonal con  $v(\mu_{ij})$  como elemento  $j$ -ésimo de la diagonal. Bajo condiciones de regularidad se tiene el siguiente teorema.

**Teorema 3.1** *El estimador  $\hat{\beta}$  de  $\beta$  es consistente y  $\sqrt{n}(\hat{\beta} - \beta)$  es asintóticamente normal cuando  $n \rightarrow \infty$  con media cero y matriz de covarianzas*

$$\mathbf{V}_\beta = \lim_{n \rightarrow \infty} n \left( \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \mathbf{A}_i \frac{\partial \mu_i}{\partial \beta} \right)^{-1} \left\{ \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \text{Cov}(\mathbf{Y}_i) \frac{\partial \mu_i}{\partial \beta} \right\} \left( \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \mathbf{A}_i \frac{\partial \mu_i}{\partial \beta} \right)^{-1}$$

donde los cálculos de los momentos para los  $\mathbf{Y}_i$ 's se hacen con respecto al verdadero modelo subyacente.

Se puede obtener un estimador consistente de la matriz de covarianzas  $\mathbf{V}_\beta$  como

$$\left[ \left( \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \mathbf{A}_i \frac{\partial \mu_i}{\partial \beta} \right)^{-1} \left\{ \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} (\mathbf{Y}_i - \mu_i(\beta)) (\mathbf{Y}_i - \mu_i(\beta))^T \frac{\partial \mu_i}{\partial \beta} \right\} \left( \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \mathbf{A}_i \frac{\partial \mu_i}{\partial \beta} \right)^{-1} \right]_{\beta=\hat{\beta}}$$

Tanto el estimador  $\hat{\beta}$  como  $\text{var}(\hat{\beta})$  son consistentes dando una correcta especificación de la regresión. La principal desventaja de  $\hat{\beta}$  es que la pérdida de eficiencia de este estimador puede ser alta si la autocorrelación entre los datos es grande. Por ello, en la siguiente sección se propone las ecuaciones de estimación generalizadas que lleva a estimadores con mayor eficiencia.

## 3.2. Metodología de las Ecuaciones de Estimación Generalizadas

Los MLG introducidos en la sección anterior se basan en la suposición de que las observaciones son independientes. Sin embargo, hay muchas situaciones en la que esto no es así sino que existe una correlación entre los datos.

La metodología EEG, desarrollada por Liang y Zeger [61], modeliza una función conocida de la esperanza marginal como una función lineal de una o varias variables explicativas, con EEG se describe la componente aleatoria del modelo para cada respuesta

marginal con una función enlace y varianzas comunes, de forma análoga a como se hace con un modelo MLG. Sin embargo, a diferencia del MLG, esta metodología tiene en cuenta una estructura para la covarianza de las medidas correlacionadas.

El método EEG proporciona un estimador consistente de los coeficientes de regresión y sus varianzas bajo hipótesis débiles sobre la correlación real entre los individuos observados. Esta aproximación evita la introducción de distribuciones multivariantes suponiendo únicamente una forma funcional para la distribución marginal en cada punto en el tiempo. La estructura de la covarianza a través del tiempo es tratada como un parámetro de ruido. El método tiene en cuenta la dependencia entre los individuos para estimar consistentemente la varianza de los estimadores propuestos incluso cuando hemos supuesto una estructura de la matriz de correlación de trabajo incorrecta.

En un modelo marginal, el efecto de la covarianza sobre la respuesta se modeliza separadamente de la correlación dentro del individuo (unidad). Debido a que el interés está en modelizar la esperanza marginal o la respuesta promedio, es natural poner más énfasis en especificar correctamente la estructura de la media marginal que en especificar la estructura de covarianza. Por eso se supone que dicha estructura tiene una forma arbitraria y se la denomina “covarianza de trabajo”.

Los datos correlacionados son modelados usando la misma función de enlace y un predictor lineal establecido como en el caso independiente. Sin embargo, Liang y Zeger [61] proponen una clase de Ecuaciones de Estimación Generalizadas (EEG) que tienen en cuenta la correlación para incrementar la eficiencia del estimador. Estas ecuaciones de estimación para estimar  $\beta$ , que son una extensión de las ecuaciones de estimación del MLG, son las siguientes:

$$\sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mu_i(\beta)) = \mathbf{0} \quad (3.1)$$

donde

$$\mathbf{V}_i = \gamma \mathbf{A}_i^{1/2} \mathbf{R}_i(\alpha) \mathbf{A}_i^{1/2}$$

siendo  $\mathbf{R}_i(\alpha)$  la matriz de correlación de trabajo. El elemento  $(j, k)$  de  $\mathbf{R}_i(\alpha)$  es la correlación entre  $y_{ij}$  y  $y_{ik}$  que es conocido, fijado o estimado. La matriz de correlación de trabajo puede depender de un vector de parámetros desconocidos  $\alpha$ , que es el mismo para todos los individuos y que deben ser estimados a partir de los datos.

Sea  $\hat{\beta}_{EEG}$  la solución final de (3.1). El teorema siguiente establece las propiedades asintóticas de  $\hat{\beta}_{EEG}$ .

**Teorema 3.2** *Suponiendo que:*

- (i)  $\hat{\alpha}$  es  $\sqrt{n}$ -consistente dado  $\beta$  y  $\gamma$ ;
  - (ii)  $\hat{\gamma}$  es  $\sqrt{n}$ -consistente dado  $\beta$ ; y
  - (iii)  $|\partial\hat{\alpha}(\beta, \gamma)/\partial\gamma| \leq H(Y, \beta)$  el cual es  $O_p(1)$ ,
- entonces  $\sqrt{n}(\hat{\beta}_{EEG} - \beta)$  es asintóticamente normal multivariante con vector de medias cero y matriz de covarianzas dada por

$$\begin{aligned} \mathbf{V}_{\hat{\beta}_{EEG}} &= \lim_{n \rightarrow \infty} n \left( \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \mathbf{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right)^{-1} \left\{ \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \mathbf{V}_i^{-1} \text{Cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right\} \\ &\quad \times \left( \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \mathbf{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right)^{-1}. \end{aligned}$$

**Demostración.** Puede verse en Liang y Zeger [61]. ■

La estimación de la varianza  $\hat{\mathbf{V}}_{\hat{\beta}_{EEG}}$  se puede obtener reemplazando  $\text{Cov}(\mathbf{Y}_i)$  por  $(\mathbf{Y}_i - \mu_i(\beta))(\mathbf{Y}_i - \mu_i(\beta))^T$  y  $\beta$ ,  $\gamma$  y  $\alpha$  por sus estimaciones en la expresión  $\mathbf{V}_{\hat{\beta}_{EEG}}$ . Como en el caso de independencia, la consistencia de  $\hat{\beta}_{EEG}$  y  $\mathbf{V}_{\hat{\beta}_{EEG}}$  depende únicamente de la correcta especificación de la media, no de la correcta elección de  $\mathbf{R}_i$ . Por supuesto, las estimaciones son más eficientes cuando se elige la forma correcta de  $\mathbf{R}_i$ , o al menos una forma que esté próxima a la verdadera matriz de correlación. Estudios de simulación y ejemplos de Liang y Zeger [61], Liang, Zeger y Qaqish [62] y Fitzmaurice, Laird y Rotnitzky [42] en general corroboran esta afirmación. Liang y Zeger [61] y Liang et al. [62] advierten que cuando las correlaciones entre los pares de respuestas son pequeños (menores que 0.4) entonces las EEG basadas en independencia obtienen estimaciones bastante eficientes.

Para llevar a cabo contrastes de hipótesis y construir intervalos de confianza, es necesario obtener los errores estándar asociados a los coeficientes de regresión  $\beta$ . Estos errores estándar se obtienen como la raíz cuadrada de los elementos de la diagonal de la matriz  $\hat{\mathbf{V}}_{\hat{\beta}_{EEG}}$ . Hay dos formas de conseguir esta matriz. Utilizando el estimador basado

en el modelo de la matriz de covarianzas para  $\hat{\beta}_{EEG}$  que es la inversa de la matriz de información observada

$$\sum_m \left( \hat{\beta}_{EEG} \right) = \mathbf{I}_0^{-1}$$

donde

$$\mathbf{I}_0 = \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \mathbf{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta}.$$

Este estimador es consistente si el modelo y la matriz de correlación de trabajo son especificados correctamente. Su uso suele ser preferible en aquellas situaciones donde tenemos un número moderado de clusters grandes (Albert y McShane [5]).

Sin embargo, en la práctica es difícil conseguir la matriz de trabajo de correlación correcta (Zorn [119]) por lo que se suele utilizar el estimador de sándwich empírico (robusto) ya que  $\sum_m \left( \hat{\beta}_{EEG} \right)$  suele subestimar el error estándar de  $\hat{\beta}_{EEG}$ . Es decir, la segunda forma es utilizar el estimador sandwich empírico (robusto) de  $Cov \left( \hat{\beta}_{EEG} \right)$  que viene dado por

$$\sum_e = \mathbf{I}_0^{-1} \mathbf{I}_1 \mathbf{I}_0^{-1}$$

donde

$$\mathbf{I}_1 = \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} \mathbf{V}_i^{-1} Cov(\mathbf{Y}_i) \mathbf{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta}$$

y la  $Cov(\mathbf{Y}_i)$  es estimada por

$$\left( \mathbf{Y}_i - \mu_i \left( \hat{\beta}_{EEG} \right) \right) \left( \mathbf{Y}_i - \mu_i \left( \hat{\beta}_{EEG} \right) \right)^T.$$

Este es un estimador consistente incluso cuando  $Var(y_{ij}) \neq v(\mu_{ij})$  o cuando  $\mathbf{R}_i(\alpha)$  no es la matriz de correlación de  $\mathbf{Y}_i$  o cuando la verdadera correlación varía a través de los clusters. Se pierde eficiencia con una mala especificación, pero si la estructura de la correlación de trabajo es más o menos correcta, se espera que la eficiencia asintótica sea relativamente alta.

### 3.2.1. Elección de la matriz de correlación de trabajo

En primer lugar, cuando el número de individuos es grande en relación al número de observaciones por individuo, la influencia de la correlación es por lo general lo suficientemente pequeña como para que los coeficientes de regresión del MLG sean casi eficientes.



La correlación entre medidas repetidas, sin embargo, puede tener un efecto importante en la varianza estimada de los coeficientes de regresión y por lo tanto tendrá que ser tomada en cuenta para hacer inferencias correctas. Puesto que es raro que la verdadera correlación sea conocida, se considera una matriz de correlación de trabajo,  $\mathbf{R}$ . Esta matriz es de tamaño  $t \times t$  porque se supone hay un número fijo de puntos en el tiempo en los que los individuos son observados. Un individuo no tiene por qué ser observado en todos los tiempos  $t$ ; la matriz de correlación de cada individuo  $\mathbf{R}_i$  es de tamaño  $t_i \times t_i$  con filas y columnas eliminadas adecuadamente si  $t_i < t$ . Se supone que la matriz de correlación  $\mathbf{R}_i$  depende de un vector de parámetros de asociación, denotado por  $\alpha$ . Es decir, que la matriz de correlación de trabajo queda completamente especificada por el vector de parámetros desconocidos  $\alpha$ . Este vector desconocido de parámetros tiene una estructura que será determinada por el investigador. No hay mucha información sobre como elegir la mejor estructura de correlación (Dahmen y Ziegler [35]) y con frecuencia es difícil de determinar. Sin embargo, la posible pérdida de eficiencia se reduce cuando crece el número de individuos (Puride et al. [94]).

Las estructuras usuales para modelizar la matriz de correlación de trabajo son la fija, independiente, intercambiable, desestructurada y autorregresiva. El amplio abanico de opciones disponible para especificar la estructura de correlación es una de las ventajas de la metodología EEG. Estas estructuras se examinan con más detalle a continuación y además aparecen resumidas en la Tabla 3.2.

Tabla 3.2

Estructura de la Matriz de Correlación de Trabajo	Estimador
<p><b>Fijada</b></p> <p><math>Corr(Y_{ij}, Y_{ik}) = r_{jk}</math>, Se tienen las matrices de correlación fijas pues hemos determinado la forma en un análisis previo.</p> <p><b>Independiente</b></p> $Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$ <p><b>m-dependiente</b></p> $Corr(Y_{ij}, Y_{i,j+s}) = \begin{cases} 1 & s = 0 \\ \alpha_s & s = 1, 2, \dots, m \\ 0 & s > m \end{cases}$ <p><b>Intercambiable</b></p> $Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha & j \neq k \end{cases}$ <p><b>Desestructurada</b></p> $Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha_{jk} & j \neq k \end{cases}$ <p><b>Auto-regresiva AR(1)</b></p> <p><math>Corr(Y_{ij}, Y_{i,j+s}) = \alpha^s</math> para <math>s = 0, 1, 2, \dots, t_i - j</math></p>	<p>La matriz de correlación no es estimada en este caso.</p> <p>La matriz de correlación no es estimada en este caso.</p> $\hat{\alpha}_s = \frac{1}{(K_s - p)} \sum_{i=1}^n \sum_{j \leq t_i - s} r_{ij} r_{i,j+s}$ $K_s = \sum_{i=1}^n (t_i - s)$ <p><math>p \Rightarrow</math> n° parámetros de Regresión</p> $\hat{\alpha} = \frac{1}{(N^* - p)\gamma} \sum_{i=1}^n \sum_{j < k} r_{ij} r_{ik}$ $N^* = 0.5 \sum_{i=1}^n t_i (t_i - 1)$ $\hat{\alpha}_{jk} = \frac{1}{(n - p)\gamma} \sum_{i=1}^n r_{ij} r_{ik}$ $\hat{\alpha} = \frac{1}{(K_1 - p)\gamma} \sum_{i=1}^n \sum_{j \leq t_i - 1} r_{ij} r_{i,j+1}$ $K_1 = \sum_{i=1}^n (t_i - 1)$

**Fijada:** Se tienen las matrices de correlación fijas cuando hemos determinado la forma en un análisis previo (Hardin y Hilbe [48]). Con esta estructura, la matriz de correlación no es estimada en cada paso, sino que la correlación se considera fija a lo largo del proceso entero.

**Independiente:** La estructura de independencia es la más simple que la matriz de correlación de trabajo puede tomar, ya que se supone que realmente no existe correlación y las observaciones son independientes. Debido a que el investigador supone que las respuestas para cada individuo son independientes entre ellas, esta aproximación sacrifica uno de los beneficios de las EEG que es no contar con la correlación para las medidas de cada individuo (Ballinger [14]). En general, esta estructura no tiene sentido lógico para datos longitudinales porque usualmente son datos altamente correlacionados. Con esta estructura, la matriz de correlación de trabajo se convierte en la matriz identidad, y las resultantes EEG son llamadas Ecuaciones de Estimación Independientes (Dahmen y Ziegler [35]). En este caso, resolver el EEG es lo mismo que ajustar el modelo de regresión usual para datos independientes, donde los estimadores paramétricos resultantes son los mismos pero sus errores estándar son diferentes.

**m-dependiente:** Con la estructura m-dependiente, las correlaciones dependen de las distancias entre las medidas, con el tiempo, tienden hacia cero para  $s \geq m$ .

**Intercambiable:** La estructura intercambiable considera varianzas iguales en la diagonal principal e iguales covarianzas en los demás elementos de la matriz, es decir, asume una misma correlación entre observaciones (Hedeker [52]), independientemente de la distancia entre tiempos de medición. Es habitual en situaciones en las que las medidas repetidas no se obtienen en función del tiempo y también cuando hay unas pocas medidas repetidas. Cuando tenemos un número arbitrario de observaciones por individuo la estructura de correlación de trabajo puede ser la independiente o la intercambiable.

**Desestructurada:** Si se dispone de suficiente cantidad de datos y los tiempos de observación son los mismos para todas las unidades se puede dejar  $\mathbf{R}(\boldsymbol{\alpha})$  sin especi-

ficar, lo que nos lleva a un modelo con estructura de correlación arbitraria y hay  $t_i(t_i - 1)/2$  parámetros a estimar. Esta estructura proporciona el estimador más eficiente para  $\beta$ , pero suele ser útil únicamente cuando no hay observaciones en muchos puntos a lo largo del tiempo (Horton y Lipsitz [53]).

**Auto-regresiva (AR-1):** Si las observaciones repetidas para cada unidad están correlacionadas de forma similar a la de un proceso Gaussiano autorregresivo, se obtiene una estructura auto-regresiva. Con una estructura de correlación auto-regresiva, las correlaciones también dependen de la distancia entre las medidas; la cual disminuye cuando aumenta la distancia.

Una forma de estimar los parámetros de correlación  $\alpha$  es utilizando el método de los momentos basados en los residuos de Pearson, dados por

$$r_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{v(\hat{\mu}_{ij})}}.$$

Las expresiones para estimar estos parámetros de correlación se pueden ver en la columna derecha de la Tabla 3.2 que dependen del parámetro de escala  $\gamma$  que se puede estimar mediante

$$\hat{\gamma} = \frac{1}{N - p} \sum_{i=1}^n \sum_{j=1}^{t_i} r_{ij}^2$$

con

$$N = \sum_{i=1}^n t_i.$$

### 3.2.2. Algoritmo para encontrar la solución EEG

Encontrar la solución EEG necesita un número de pasos, que incluye la especificación del modelo marginal para el primer momento, especificar la función varianza para la relación entre el primer y segundo momento, elegir una matriz de correlación de trabajo, calcular un estimador inicial de  $\beta$ , y entonces utilizar este estimador en un proceso de estimación iterativo. En detalle:

El **primer paso** del método EEG es relacionar la respuesta marginal  $\mu_{ij} = E(y_{ij})$  a una combinación lineal de variables explicativas:  $g(\mu_{ij}) = \mathbf{x}_{ij}^T \beta$ , donde  $\beta = (\beta_1, \dots, \beta_p)$

es un vector  $p \times 1$  de parámetros desconocidos y  $g$  es una función de enlace conocida. La función de enlace natural para respuestas binarias es la función logit  $g(x) = \log\left(\frac{x}{1-x}\right)$  y para respuestas de conteo de Poisson es la función logaritmo  $g(x) = \log(x)$ .

El **segundo paso** es describir la varianza de  $y_{ij}$  como una función de la media:  $Var(y_{ij}) = v(\mu_{ij})\gamma$ , donde  $v$  es una función varianza conocida y  $\gamma$  es un posible parámetro de escala desconocido. Para respuestas binarias,  $v(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$ ; y para respuestas Poisson,  $v(\mu_{ij}) = \mu_{ij}$  y  $\gamma = 1$ . Sobredispersión ( $\gamma > 1$ ) puede existir para la binomial o datos de conteo, pero el uso de la matriz de covarianza empírica para el procedimiento EEG es robusto a esta sobredispersión.

El **tercer paso** es elegir la forma de una matriz  $t_i \times t_i$  de correlación  $\mathbf{R}_i(\boldsymbol{\alpha})$  para cada  $\mathbf{Y}_i = (y_{i1}, \dots, y_{it_i})^T$ . El elemento  $(j, k)$  de  $\mathbf{R}_i(\boldsymbol{\alpha})$  es la correlación conocida, hipotética o estimada entre  $y_{ij}$  e  $y_{ik}$ . Esta matriz de correlación puede depender de un vector de parámetro desconocido  $\boldsymbol{\alpha}$ , que es el mismo para todos los individuos. Suponemos que  $\mathbf{R}_i(\boldsymbol{\alpha})$  es conocida excepto para un número fijo de parámetros  $\boldsymbol{\alpha}$  que debe ser estimado de los datos. Aunque esta matriz de correlación puede ser diferente de sujeto a sujeto, comúnmente utilizamos una matriz de correlación  $\mathbf{R}(\boldsymbol{\alpha})$  que aproxima la dependencia media entre las observaciones repetidas sobre los sujetos.

El método EEG da estimadores consistentes de los coeficientes de la regresión y de sus varianzas, incluso especificando mal la estructura de la matriz de correlación de trabajo. Además, la pérdida de eficiencia de una elección incorrecta de  $\mathbf{R}$  no tiene consecuencias cuando el número de individuos es grande.

El **cuarto paso** del método EEG es estimar el vector de parámetros  $\boldsymbol{\beta}$  y su matriz de covarianza. Primero, sea  $\mathbf{A}_i$  la matriz diagonal  $t_i \times t_i$  cuyo elemento  $j$  -ésimo de la diagonal es  $v(\mu_{ij})$ . La matriz de covarianza para  $\mathbf{Y}_i$  es  $\mathbf{V}_i(\boldsymbol{\alpha}) = \gamma \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$ . La estimación EEG de  $\boldsymbol{\beta}$  es la solución de la ecuación de estimación

$$\mathbf{U}(\boldsymbol{\theta}) = \sum_{i=1}^n \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^T [\mathbf{V}_i(\hat{\boldsymbol{\alpha}})]^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

donde  $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{it_i})^T$ ,  $\mathbf{0}$  es el vector  $p \times 1$   $(0, \dots, 0)^T$  y  $\hat{\boldsymbol{\alpha}}$  es un estimador consistente de  $\boldsymbol{\alpha}$ .

Describimos a continuación el algoritmo para el ajuste del modelo específico de ajuste utilizando EEG: <sup>1</sup>

1. Se calcula un estimador inicial de  $\beta$  con las EEG suponiendo independencia.
2. Se calcula la matriz de correlación de trabajo  $\mathbf{R}$  utilizando el residuo estándar de Pearson  $r_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{v(\hat{\mu}_{ij})}}$ , el actual valor de  $\beta$ , y la estructura que hayamos supuesto de  $\mathbf{R}$ .
3. Calculamos una estimación de la covarianza:

$$\hat{\mathbf{V}}_i(\hat{\alpha}) = \gamma \mathbf{A}_i^{1/2} \mathbf{R}_i(\hat{\alpha}) \mathbf{A}_i^{1/2}$$

4. Actualizamos  $\hat{\beta}$ :

$$\begin{aligned} \hat{\beta}_{r+1} = \hat{\beta}_r + & \left[ \sum_{i=1}^n \mathbf{X}_i^T \mathbf{\Lambda}_i \mathbf{V}_i^{-1}(\hat{\alpha}) \mathbf{\Lambda}_i \mathbf{X}_i \right]^{-1} \\ & \times \left[ \sum_{i=1}^n \mathbf{X}_i^T \mathbf{\Lambda}_i \mathbf{V}_i^{-1}(\hat{\alpha}) (\mathbf{Y}_i - \boldsymbol{\mu}_i) \right] \end{aligned}$$

donde

$$\mathbf{\Lambda}_i = \text{diag} \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \right).$$

5. Iteramos 2-4 hasta que se de la convergencia.

### 3.3. Procedimientos alternativos a las EEG

El método de Liang y Zeger presenta un método de estimación en dos pasos. El primero es un método de cuasi-verosimilitud para estimar los parámetros de regresión que caracterizan la dependencia de las variables explicativas. El segundo es el método de los momentos robusto para la estimación de los parámetros de correlación que incorpora la dependencia entre las respuestas. Este método necesita varias iteraciones entre estos dos pasos hasta que converja. Para la estimación de los parámetros de correlación se suele

<sup>1</sup>Nótese que este no es en general un método de estimación basado en la verosimilitud.

utilizar los residuos de Pearson que se supone implícitamente que se distribuyen normalmente. Sin embargo, esta normalidad de los residuos de Pearson, no se da necesariamente cuando las respuestas son discretas como en el caso de respuestas binarias.

En los modelos lineales generalizados univariantes, se han desarrollado otros residuos además de los de Pearson que también son utilizados con frecuencia como el de Anscombe, el de desviación o los residuos de  $\phi$ -divergencia. Park et al. [84] consideran como alternativa a los residuos de Pearson los residuos de Anscombe y de desviación. En esta sección, se extiende el procedimiento EEG utilizando los residuos  $\phi$ -divergencia que contienen como caso particular al de Pearson y al desviación para estimar los parámetros de correlación.

Desde que Liang y Zeger unificaron criterios para el análisis de datos longitudinales, han sido muchas las modificaciones y las extensiones. Prentice [92] propuso una modelización conjunta para la media y para los parámetros de correlación. En vez de los parámetros de correlación, Lipsitz et al. [64] propusieron modelar la dependencia entre las respuestas binarias utilizando los “odds ratios”. Enfoques parecidos han sido propuestos por Liang et al. [62] y Paik [87]. Zhao y Prentice [118] propusieron una nueva clase de distribuciones multivariantes llamadas distribuciones exponenciales parcialmente. Si la estructura de la covarianza está correctamente especificada, las EEG derivadas por Liang y Zeger vuelven a ser ecuaciones de verosimilitud para la nueva densidad. Todos estos enfoques introducen simultáneamente la estimación en las EEG de los parámetros de regresión y de los parámetros de correlación. Como en Liang y Zeger [61] el segundo paso está basado en el residuo de Pearson, por lo que igualmente se podrían extender utilizando los residuos  $\phi$ -divergencia.

El objetivo es estudiar si mejora el ajuste al estimar los parámetros de correlación en lugar de con el residuo de Pearson con otros residuos. Para ello, vamos a considerar los residuos phi-divergencia como generalización tanto del residuo de Pearson como del de desviación como desarrollamos a continuación.

### 3.3.1. Nuevos procedimientos basados en medidas de divergencia

Las EEG están definidas por

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^n \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^T [\mathbf{V}_i(\hat{\boldsymbol{\alpha}})]^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0} \quad (3.2)$$

donde  $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{it_i})^T$ ,  $\mathbf{0} = (0, \dots, 0)^T$  y  $\hat{\boldsymbol{\alpha}}$  es un estimador consistente de  $\boldsymbol{\alpha}$ .

El estimador de  $\boldsymbol{\beta}$  se obtiene como solución de (3.2). En general, la estimación necesita repeticiones entre este método de quasi-verosimilitud para  $\boldsymbol{\beta}$  y un método robusto para estimar  $\boldsymbol{\alpha}$  y  $\gamma$ . Sea  $\hat{\boldsymbol{\beta}}_{EG}$  la solución final de (3.2). Bajo condiciones de regularidad, el Teorema 3.2 muestra que  $n^{1/2} (\hat{\boldsymbol{\beta}}_{EG} - \boldsymbol{\beta})$  es asintóticamente normal multivariante de media cero y matriz de covarianza  $\mathbf{V}_{\hat{\boldsymbol{\beta}}_{EG}}$ . Obsérvese que la varianza asintótica de  $\hat{\boldsymbol{\beta}}_{EG}$  no depende de la elección del estimador para  $\boldsymbol{\alpha}$  y  $\gamma$  siempre y cuando sean estimadores  $\sqrt{n}$ -consistentes.

Liang y Zeger [61] utilizaron para estimar el parámetro  $\boldsymbol{\alpha}$  y  $\gamma$  los residuos de Pearson definidos por

$$r_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{v(\hat{\mu}_{ij})}} \quad (3.3)$$

que reciben este nombre porque el estadístico de Pearson viene dado por

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^{t_i} r_{ij}^2.$$

Sin embargo, éstos no son los únicos residuos, Park et al. [84] proponen utilizar el residuo de desviación definido por

$$d_{ij} = \text{sig}(y_{ij} - \hat{\mu}_{ij}) e_{ij} \quad (3.4)$$

que es la raíz cuadrada con signo de las componentes del test del cociente de verosimilitud que viene dado por

$$D = \sum_{i=1}^n \sum_{j=1}^{t_i} e_{ij}^2$$

o el residuo Anscombe (Anscombe [8]).

En el caso de variables binarias, el residuo de Pearson y el de desviación son respectivamente:

$$r_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1 - \hat{\mu}_{ij})}}$$



y

$$d_{ij} = \text{sig}(y_{ij} - \hat{\mu}_{ij}) \sqrt{2 \left\{ y_{ij} \log \frac{y_{ij}}{\hat{\mu}_{ij}} + (1 - y_{ij}) \log \frac{1 - y_{ij}}{1 - \hat{\mu}_{ij}} \right\}}.$$

Pardo et al. [81] proponen la familia de residuos de  $\phi$ -divergencia, para detectar “outliers” en regresión logística, y viene dada por

$$c_{ij}^{\phi} = \text{sig}(y_{ij} - \hat{\mu}_{ij}) \sqrt{\frac{2}{\phi''(1)} \left\{ \hat{\mu}_{ij} \phi\left(\frac{y_{ij}}{\hat{\mu}_{ij}}\right) + (1 - \hat{\mu}_{ij}) \phi\left(\frac{1 - y_{ij}}{1 - \hat{\mu}_{ij}}\right) \right\}^{1/2}} \quad (3.5)$$

con  $\phi \in \Phi^*$  la clase de todas las funciones convexas  $\phi(x)$ ,  $x > 0$ , tales que en  $x = 1$ ,  $\phi(1) = \phi'(1) = 0$ ,  $\phi''(1) > 0$ , y en  $x = 0$ ,  $0\phi\left(\frac{0}{0}\right) = 0$  y  $0\phi\left(\frac{p}{0}\right) = p \lim_{n \rightarrow \infty} \frac{\phi(u)}{u}$  que contiene como casos particulares a los residuos de Pearson y de desviación considerando  $\phi(x) = \frac{1}{2}(x-1)^2$  y  $\phi(x) = x \log x - x + 1$ , respectivamente.

En el caso de variables de Poisson, el residuo de Pearson, el de desviación y el de  $\phi$ -divergencia son

$$\begin{aligned} r_{ij} &= \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}, \\ d_{ij} &= \text{sig}(y_{ij} - \hat{\mu}_{ij}) \sqrt{2 \left\{ y_{ij} \log \frac{y_{ij}}{\hat{\mu}_{ij}} - y_{ij} + \hat{\mu}_{ij} \right\}}, \\ c_{ij}^{\phi} &= \text{sig}(y_{ij} - \hat{\mu}_{ij}) \sqrt{\frac{2}{\phi''(1)} \left\{ \hat{\mu}_{ij} \phi\left(\frac{y_{ij}}{\hat{\mu}_{ij}}\right) \right\}^{1/2}} \end{aligned} \quad (3.6)$$

con  $\phi \in \Phi^*$  la clase de todas las funciones convexas  $\phi(x)$ ,  $x > 0$ , tales que en  $x = 1$ ,  $\phi(1) = \phi'(1) = 0$ ,  $\phi''(1) > 0$ , y en  $x = 0$ ,  $0\phi\left(\frac{0}{0}\right) = 0$  y  $0\phi\left(\frac{p}{0}\right) = p \lim_{n \rightarrow \infty} \frac{\phi(u)}{u}$  que contiene como casos particulares a los residuos de Pearson y de desviación considerando  $\phi(x) = \frac{1}{2}(x-1)^2$  y  $\phi(x) = x \log x - x + 1$ , respectivamente.

Por tanto, si los estimadores de  $\alpha$  y  $\gamma$  que resultan al utilizar los residuos de  $\phi$ -divergencia en lugar de los residuos de Pearson son  $\sqrt{n}$ -consistentes las propiedades asintóticas de  $\hat{\beta}_{EEG}$  se conservan. La consistencia de los estimadores de  $\alpha$  y  $\gamma$  cuando se utilizan los residuos de Pearson o de desviación pueden verse en McCullagh y Nelder [68]. En el teorema siguiente se establece la consistencia de  $\alpha$  y  $\gamma$  cuando se utilizan los residuos de  $\phi$ -divergencia.

**Teorema 3.3** *Bajo condiciones de regularidad dadas en Sección 3.3 de Liang y Zeger [61] y para  $\phi \in \Phi^*$ , se tiene que:*

(i)

$$\hat{\gamma}_\phi = \frac{1}{N-p} \sum_{i=1}^n \sum_{j=1}^{t_i} \left( c_{ij}^\phi \right)^2,$$

es  $\sqrt{n}$ -consistente dado  $\beta$ .

(ii)

$$\hat{\alpha}_\phi = \frac{1}{N-p} \sum_{i=1}^n \sum_{j < k} c_{ij}^\phi c_{ik}^\phi \quad \text{con } N = \sum_{i=1}^n t_i$$

(estructura de la matriz correlación desestructurada)

es  $\sqrt{n}$ -consistente dado  $\beta$  y  $\gamma$ ;

**Demostración.** Se considera la función

$$f(x) = \frac{2}{\phi''(1)} \left\{ \hat{\mu}_{ij} \phi \left( \frac{y_{ij}}{\hat{\mu}_{ij}} \right) + (1 - \hat{\mu}_{ij}) \phi \left( \frac{1 - y_{ij}}{1 - \hat{\mu}_{ij}} \right) \right\}$$

El desarrollo en serie de Taylor de segundo orden de esta función alrededor de  $\hat{\mu}_{ij}$ , viene dado por

$$f(x) = f(\hat{\mu}_{ij}) + f'(\hat{\mu}_{ij})(x - \hat{\mu}_{ij}) + \frac{1}{2} f''(\hat{\mu}_{ij})(x - \hat{\mu}_{ij})^2 + o\left((x - \hat{\mu}_{ij})^2\right).$$

Es inmediato que

$$f(\hat{\mu}_{ij}) = 0, \quad f'(\hat{\mu}_{ij}) = 0$$

y

$$f''(\hat{\mu}_{ij}) = \frac{2}{\hat{\mu}_{ij}(1 - \hat{\mu}_{ij})}.$$

Entonces,

$$\left( c_{ij}^\phi \right)^2 = \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}(1 - \hat{\mu}_{ij})} + o\left((y_{ij} - \hat{\mu}_{ij})^2\right)$$

y por tanto  $c_{ij}^\phi$  y  $r_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1 - \hat{\mu}_{ij})}}$  son variables aleatorias con las mismas propiedades asintóticas.

Puesto que  $\hat{\gamma}_\phi$  y  $\hat{\alpha}_\phi$  para cualquier estructura de la matriz de correlación son combinaciones de  $c_{ij}^\phi$ . Y los valores de  $\hat{\gamma}$  y  $\hat{\alpha}$  para cualquier estructura de la matriz de correlación son combinaciones de  $r_{ij}$ . Al ser  $\hat{\gamma}$  y  $\hat{\alpha}$   $\sqrt{n}$ -consistentes, se tiene que  $\hat{\gamma}_\phi$  y  $\hat{\alpha}_\phi$  son también  $\sqrt{n}$ -consistente para variables respuesta binaria.

De forma análoga se puede demostrar para variables respuesta de Poisson. ■

De esta forma, al igual que Park et al. [84], consideraron en lugar de los residuos de Pearson, dos residuos alternativos (Residuos de desviación y de Anscombe) para estimar los parámetros de correlación. En este capítulo, se propone la familia de residuos  $\phi$ -divergencia que contiene a los residuos de Pearson y de desviación para sustituir a los residuos de Pearson en el procedimiento clásico EEG y estudiar la nueva familia de procedimientos alternativa.

El nuevo procedimiento sigue los pasos:

1. Se calcula un estimador inicial de  $\beta$  con las EEG suponiendo independencia.
2. Se calcula la matriz de correlación de trabajo  $\mathbf{R}$  utilizando la familia de residuos de  $\phi$ -divergencia  $c_{ij}^\phi$  definidos anteriormente. Además del residuo necesitamos el actual valor de  $\beta$  y la estructura que hayamos supuesto de  $\mathbf{R}$ .
3. Calculamos una estimación de la covarianza:

$$\mathbf{V}_i(\hat{\alpha}) = \gamma \mathbf{A}_i^{1/2} \mathbf{R}_i(\hat{\alpha}) \mathbf{A}_i^{1/2}$$

4. Actualizamos  $\hat{\beta}$  :

$$\begin{aligned} \hat{\beta}_{r+1} = \hat{\beta}_r + & \left[ \sum_{i=1}^n \mathbf{X}_i^T \mathbf{\Lambda}_i \mathbf{V}_i^{-1}(\hat{\alpha}) \mathbf{\Lambda}_i \mathbf{X}_i \right]^{-1} \\ & \times \left[ \sum_{i=1}^n \mathbf{X}_i^T \mathbf{\Lambda}_i \mathbf{V}_i^{-1}(\hat{\alpha}) (\mathbf{Y}_i - \mu_i) \right] \end{aligned}$$

donde

$$\mathbf{\Lambda}_i = \text{diag} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)$$

5. Volvemos al paso 2 hasta que converja.

### 3.4. Ejemplos

En este apartado se implementan los nuevos procedimientos de estimación alternativos a las EEG, introducidos en la sección anterior, con programas en SAS. El proce-

dimiento estándar PROC GENMOD es el que ofrece SAS para resolver las EEG. Este procedimiento es utilizado en nuestros programas suponiendo independencia para encontrar el estimador inicial de  $\beta$  con el que empezar el algoritmo.

Estos nuevos procedimientos dependen de la familia  $\phi$  que debemos fijar en un ejemplo práctico. Una familia paramétrica  $\phi$  muy conocida y estudiada es la introducida por Cressie y Read [33] que viene dada por

$$\phi_{(\lambda)}(x) = (\lambda(\lambda + 1))^{-1} (x^{\lambda+1} - x - \lambda(x - 1)), \lambda \neq 0, \lambda \neq -1 \quad (3.7)$$

$$\phi_{(0)}(x) = \lim_{\lambda \rightarrow 0} \phi_{(\lambda)}(x) = x \log x - x + 1$$

$$\phi_{(-1)}(x) = \lim_{\lambda \rightarrow -1} \phi_{(\lambda)}(x) = -\log x + x - 1.$$

En realidad contiene los residuos clásicos de Pearson ( $\lambda = 1$ ) y de desviación ( $\lambda = 0$ ) y otros nuevos, correspondientes a los diferentes valores de  $\lambda$ .

Se consideran en todos los ejemplos los procedimientos correspondientes a los residuos  $\phi$ -divergencia,  $c_{ij}^{\phi(\lambda)}$  con  $\lambda = -1/2, 0, 2/3, 1, 2, 3$ . Se muestran las estimaciones de los parámetros del modelo para estos diferentes procedimientos así como sus errores estándar.

Se ilustran los nuevos procedimientos con 3 ejemplos que se corresponden con diferentes situaciones en las que surgen datos longitudinales.

### 3.4.1. Ejemplo 1. Tabla de Contingencia

Las tablas de contingencia proporcionan una forma de representar medidas repetidas. En una tabla de contingencia, los individuos son sus propios controles y reciben dos o más tratamientos o condiciones en dos o más periodos consecutivos. Podemos usar el método EEG para analizar tales datos, manejando los individuos como clusters y considerando el tratamiento como una variable dependiente del tiempo.

Los siguientes datos proceden de un estudio cruzado de dos periodos en el que se estudian tres tratamientos.

		Respuestas				
Edad	Secuencia	FF	FU	UF	UU	Total
viejo	A:B	12	12	6	20	50
viejo	B:P	8	5	6	31	50
viejo	P:A	5	3	22	20	50
joven	B:A	19	3	25	3	50
joven	A:P	25	6	6	13	50
joven	P:B	13	5	21	11	50

La Tabla contiene 24 datos agregados (conteos). Sin embargo el modelo se plantea para los datos desagregados  $300 \times 2 = 600$  respuestas individuales.

Esta es una tabla de contingencia donde los pacientes han sido estratificados en 2 grupos por edades y, dentro de cada grupo, han sido asignados a una de las 3 secuencias de los tratamientos. Estos datos pueden ser modelados con parámetros para efecto del periodo, efectos por *medicamento A* y *medicamento B* relativo al placebo (P), efectos de retardo para el *medicamento A* y *medicamento B*, y las interacciones de *periodo*  $\times$  *edad* y *medicamento*  $\times$  *edad*.

El siguiente DATA introduce los datos en SAS como un conjunto de datos cruzados. La variable *edad* contiene información sobre si el individuo es joven o viejo, y la variable *secuencia* contiene dos letras que describen la secuencia de tratamientos para este grupo. Por ejemplo, el valor AB significa que el tratamiento A fue recibido en el primer periodo y el tratamiento B fue recibido en el segundo, el valor BP significa que el tratamiento B se dio primero y el placebo después, etc... Las variables *tiempo1* y *tiempo2* tienen valores F y U dependiendo de si el tratamiento produce una respuesta favorable o desfavorable. Los datos son normalmente frecuencias, y la variable *count* contiene la frecuencia para cada perfil de respuesta para cada combinación de *secuencia* y *edad*. El siguiente DATA crea una observación para cada individuo.

```
data cross (drop=count);
  input edad $ secuencia $ tiempo1 $ tiempo2 $ count;
```

```

do i=1 to count;
  output;
end;
datalines;
viejo AB F F 12
viejo AB F U 12
viejo AB U F 6
viejo AB U U 20
viejo BP F F 8
viejo BP F U 5
viejo BP U F 6
viejo BP U U 31
viejo PA F U 3
viejo PA U F 22
viejo PA U U 20
joven BA F F 19
joven BA F U 3
joven BA U F 25
joven BA U U 3
joven AP F F 25
joven AP F U 6
joven AP U F 6
joven AP U U 13
joven PB F F 13
joven PB F U 5
joven PB U F 21
joven PB U U 11
;

```

El siguiente DATA crea una observación para cada respuesta en cada periodo de forma que los datos tengan una estructura correcta para el análisis EEG. La variable

*periodo* es una variable que indica si la observación es del primer periodo. La variable *respuesta* contiene el valor 1 si la respuesta fue favorable y 0 si no lo fue.

```
data cross2;
  set cross;
  subject=_n_;
  periodo=1;
  tratamiento = substr(secuencia, 1, 1);
  carry='N';
  respuesta = tiempo1;
  output;
  periodo=0;
  tratamiento = substr(secuencia, 2, 1);
  carry = substr(secuencia, 1, 1);
  if carry='P' then carry='N';
  respuesta = tiempo2;
  output;
  run;
proc print data=cross2(obs=15);
run;
```

La variable *carry* toma el valor N (no) si la observación procede del primer periodo; toma el valor A o B si viene del segundo periodo y el tratamiento en el primer periodo es A o B, respectivamente. Si el individuo recibe el placebo en el primer periodo, el valor de *carry* es también N para la observación en el segundo periodo.

La matriz de correlación de trabajo que hemos elegido es la desestructurada pues el número de datos no es grande por lo que el coste computación no es importante y además para el caso en el de no tener muchos datos la literatura sugiere esa elección.

Ya que hay 300 individuos en la tabla de estudio, hay 300 clusters o unidades experimentales en el análisis EEG. Como tenemos respuestas para ambos periodos, el tamaño del cluster es 2. No hay valores missing, así que tanto el máximo como el mínimo cluster tienen tamaño 2. Un análisis de regresión logística es apropiado para estos datos así que especificamos DIST=BIN en la sentencia MODEL. El link logit es usado por defecto.

Utilizamos el procedimiento estándar de SAS PROC GENMOD para calcular el valor inicial de los parámetros y después utilizamos los procedimientos alternativos a las EEG.

Vamos a mostrar únicamente las estimaciones de los parámetros junto con su error estándar obtenido utilizando los nuevos procedimientos propuestos. En las Tablas 3.3 y 3.4 se muestran las estimaciones de los parámetros y sus errores estándar, respectivamente. Señalar otra vez, que al considerar la familia de residuos basada en la familia de Cressie y Read se tiene un nuevo procedimiento para cada valor de  $\lambda$ . Además, para  $\lambda = 1$  se tiene el procedimiento clásico de EEG.

Tabla 3.3. Estimadores de los parámetros

<b>Parámetros</b>	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
<i>Intercepto</i>	0.5281	0.5166	0.5129	0.5126	0.5157	0.5246
<i>periodo</i>	-1.1584	-1.1560	-1.1559	-1.1553	-1.1559	-1.1577
<i>edad_mayor</i>	-1.4970	-1.4988	-1.4993	-1.4993	-1.4990	-1.4976
<i>tratamiento_A</i>	1.2439	1.2516	1.2541	1.2543	1.2522	1.2462
<i>tratamiento_B</i>	0.3073	0.3321	0.3401	0.3405	0.3339	0.3149
<i>periodo <math>\times</math> edad_mayor</i>	0.7084	0.7087	0.7088	0.7088	0.7087	0.7084

Tabla 3.4. Errores estándar de las estimaciones

<b>Parámetros</b>	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
<i>Intercepto</i>	0.2041	0.2056	0.2062	0.2063	0.2057	0.2044
<i>periodo</i>	0.2301	0.2303	0.2304	0.2304	0.2303	0.2302
<i>edad_mayor</i>	0.2583	0.2583	0.2583	0.2583	0.2583	0.2583
<i>tratamiento_A</i>	0.2030	0.2010	0.2010	0.2011	0.2010	0.2020
<i>tratamiento_B</i>	0.2044	0.2018	0.2016	0.2016	0.2017	0.2033
<i>periodo <math>\times</math> edad_mayor</i>	0.3127	0.3130	0.3131	0.3131	0.3130	0.3128

Las estimaciones obtenidas con los diferentes procedimientos son análogas pero es un hecho a destacar que los mayores errores estándar se obtienen cuando se considera el procedimiento clásico EEG ( $\lambda = 1$ ).

Por otra parte, de los procedimientos alternativos considerados con los que se obtienen



menores errores cuadráticos medios para la mayoría de los parámetros son  $\lambda = -1/2$  y  $\lambda = 0$ .

### 3.4.2. Ejemplo 2. Insuficiencia respiratoria

El siguiente ejemplo consiste en un ensayo clínico en el que se comparan dos tratamientos distintos que tratan enfermedades respiratorias, ver Koch et al. [56]. En cada uno de los centros donde se realiza el estudio, se eligen de forma aleatoria a los pacientes a los cuales se les asigna o un tratamiento activo o un placebo. Durante el tratamiento, se hace un seguimiento a estos pacientes anotando su capacidad respiratoria en cuatro visitas, determinando Status respiratorio. El número total de pacientes estudiados fue 111 de los cuales 54 recibieron un tratamiento con un principio activo y 57 con un placebo.

El siguiente DATA introduce los datos de capacidad respiratoria y crea una observación para cada respuesta. La línea de referencia (*baseline*) y las respuestas puestas en su lugar son medidas en un escala de cinco-puntos, desde terrible a excelente, y esta respuesta ordinal es analizada después. Para el análisis, las salidas dicotómicas de si el paciente tuvo una respuesta buena o excelente es analizada con una regresión logística. El segundo DATA crea el conjunto de datos de SAS RESP2 y analiza variables respuesta *dichot* y variables baseline dicotómicas *di\_base*. Nótese que la variable baseline, fue registrada en una escala de cinco-puntos.

```
data resp;
    input centro id tratamiento $ sexo $ edad baseline
    visit1-visit4 @@;
    visit=1; outcome=visit1; output;
    visit=2; outcome=visit2; output;
    visit=3; outcome=visit3; output;
    visit=4; outcome=visit4; output;
    datalines;
    1 53 A F 32 1 2 2 4 2 2 30 A F 37 1 3 4 4 4
    1 18 A F 47 2 2 3 4 4 2 52 A F 39 2 3 4 4 4
    1 54 A M 11 4 4 4 4 2 2 23 A F 60 4 4 3 3 4
    1 12 A M 14 2 3 3 3 2 2 54 A F 63 4 4 4 4 4
```

1 51 A M 15 0 2 3 3 3 2 12 A M 13 4 4 4 4 4  
1 20 A M 20 3 3 2 3 1 2 10 A M 14 1 4 4 4 4  
1 16 A M 22 1 2 2 2 3 2 27 A M 19 3 3 2 3 3  
1 50 A M 22 2 1 3 4 4 2 16 A M 20 2 4 4 4 3  
1 3 A M 23 3 3 4 4 3 2 47 A M 20 2 1 1 0 0  
1 32 A M 23 2 3 4 4 4 2 29 A M 21 3 3 4 4 4  
1 56 A M 25 2 3 3 2 3 2 20 A M 24 4 4 4 4 4  
1 35 A M 26 1 2 2 3 2 2 2 A M 25 3 4 3 3 1  
1 26 A M 26 2 2 2 2 2 2 15 A M 25 3 4 4 3 3  
1 21 A M 26 2 4 1 4 2 2 25 A M 25 2 2 4 4 4  
1 8 A M 28 1 2 2 1 2 2 9 A M 26 2 3 4 4 4  
1 30 A M 28 0 0 1 2 1 2 49 A M 28 2 3 2 2 1  
1 33 A M 30 3 3 4 4 2 2 55 A M 31 4 4 4 4 4  
1 11 A M 30 3 4 4 4 3 2 43 A M 34 2 4 4 2 4  
1 42 A M 31 1 2 3 1 1 2 26 A M 35 4 4 4 4 4  
1 9 A M 31 3 3 4 4 4 2 14 A M 37 4 3 2 2 4  
1 37 A M 31 0 2 3 2 1 2 36 A M 41 3 4 4 3 4  
1 23 A M 32 3 4 4 3 3 2 51 A M 43 3 3 4 4 2  
1 6 A M 34 1 1 2 1 1 2 37 A M 52 1 2 1 2 2  
1 22 A M 46 4 3 4 3 4 2 19 A M 55 4 4 4 4 4  
1 24 A M 48 2 3 2 0 2 2 32 A M 55 2 2 3 3 1  
1 38 A M 50 2 2 2 2 2 2 3 A M 58 4 4 4 4 4  
1 48 A M 57 3 3 4 3 4 2 53 A M 68 2 3 3 3 4  
1 5 P F 13 4 4 4 4 4 2 28 P F 31 3 4 4 4 4  
1 19 P F 31 2 1 0 2 2 2 5 P F 32 3 2 2 3 4  
1 25 P F 35 1 0 0 0 0 2 21 P F 36 3 3 2 1 3  
1 28 P F 36 2 3 3 2 2 2 50 P F 38 1 2 0 0 0  
1 36 P F 45 2 2 2 2 1 2 1 P F 39 1 2 1 1 2  
1 43 P M 13 3 4 4 4 4 2 48 P F 39 3 2 3 0 0  
1 41 P M 14 2 2 1 2 3 2 7 P F 44 3 4 4 4 4  
1 34 P M 15 2 2 3 3 2 2 38 P F 47 2 3 3 2 3

```

1 29 P M 19 2 3 3 0 0 2 8 P F 48 2 2 1 0 0
1 15 P M 20 4 4 4 4 4 2 11 P F 48 2 2 2 2 2
1 13 P M 23 3 3 1 1 1 2 4 P F 51 3 4 2 4 4
1 27 P M 23 4 4 2 4 4 2 17 P F 58 1 4 2 2 0
1 55 P M 24 3 4 4 4 3 2 39 P M 11 3 4 4 4 4
1 17 P M 25 1 1 2 2 2 2 40 P M 14 2 1 2 3 2
1 45 P M 26 2 4 2 4 3 2 24 P M 15 3 2 2 3 3
1 40 P M 26 1 2 1 2 2 2 41 P M 15 4 3 3 3 4
1 44 P M 27 1 2 2 1 2 2 33 P M 19 4 2 2 3 3
1 49 P M 27 3 3 4 3 3 2 13 P M 20 1 4 4 4 4
1 39 P M 28 2 1 1 1 1 2 34 P M 20 3 2 4 4 4
1 2 P M 28 2 0 0 0 0 2 45 P M 33 3 3 3 2 3
1 14 P M 30 1 0 0 0 0 2 22 P M 36 2 4 3 3 4
1 10 P M 37 3 2 3 3 2 2 18 P M 38 4 3 0 0 0
1 31 P M 37 1 0 0 0 0 2 35 P M 42 3 2 2 2 2
1 7 P M 43 2 3 2 4 4 2 44 P M 43 2 1 0 0 0
1 52 P M 43 1 1 1 3 2 2 6 P M 45 3 4 2 1 2
1 4 P M 44 3 4 3 4 2 2 46 P M 48 4 4 0 0 0
1 1 P M 46 2 2 2 2 2 2 31 P M 52 2 3 4 3 4
1 46 P M 49 2 2 2 2 2 2 42 P M 66 3 3 3 4 4
1 47 P M 63 2 2 2 2 2

;
data resp2; set resp;
    dichot=(outcome=3 or outcome=4);
    di_base = (baseline=3 or baseline=4);
run;

```

Con esta sentencia estamos dicotomizando las variables *dichot* y *di\_base* de la siguiente manera,

$$dichot \begin{cases} outcome=1 \text{ ó } 2 & \Rightarrow 0 \\ outcome=3 \text{ ó } 4 & \Rightarrow 1 \end{cases}$$

$$di\_base \begin{cases} \text{outcome}=1 \text{ ó } 2 \Rightarrow 0 \\ \text{outcome}=3 \text{ ó } 4 \Rightarrow 1 \end{cases}$$

Quedando el conjunto de datos de la manera que se muestra a continuación adjuntando un trozo del mismo para el caso del *Centro1*

Tabla 3.5. Datos de insuficiencia respiratoria para 56 individuos del Centro 1.

Paciente	Tratamiento	Sexo	Edad	Status Respiratorio(0 = <i>malo</i> , 1 = <i>bueno</i> )				
				Baseline	Visit1	Visit2	Visit3	Visit4
1	P	M	46	0	0	0	0	0
2	P	M	28	0	0	0	0	0
3	A	M	23	1	1	1	1	1
4	P	M	44	1	1	1	1	0
5	P	F	13	1	1	1	1	1
6	A	M	34	0	0	0	0	0
7	P	M	43	0	1	0	1	1
8	A	M	28	0	0	0	0	0
9	A	M	31	1	1	1	1	1
10	P	M	37	1	0	1	1	0

Hacemos una regresión logística por lo que la distribución es binomial y la función de enlace considerada es la logit, la estructura de la matriz de trabajo que consideramos es la desestructurada. Utilizamos el proc estándar de SAS PROC GENMOD para calcular los valores iniciales de los parámetros y a partir de ahí utilizamos los procedimientos alternativos a las EEG.

En este ejemplo implementamos el modelo

$$\begin{aligned} \text{logit}(P(\text{dichot} = 1)) = & \beta_0 + \beta_1 \cdot \text{centro\_1} + \beta_2 \cdot \text{centro\_2} + \beta_3 \cdot \text{sexo\_F} \\ & + \beta_4 \cdot \text{sexo\_M} + \beta_5 \cdot \text{tratamiento\_A} + \beta_6 \cdot \text{tratamiento\_P} \\ & + \beta_7 \cdot \text{edad} + \beta_8 \cdot \text{di\_base} \end{aligned}$$

En la Tabla 3.6 se muestra las estimaciones de los parámetros de este modelo utilizando los diferentes procedimientos alternativos a las EEG propuestas en la sección

anterior. Cada uno de estos procedimientos consiste en considerar un miembro diferente de la familia de residuos  $c_{ij}^{\phi(\lambda)}$  para estimar el parámetro de correlación. Por tanto, cada columna de la tabla se corresponde con cada uno de los diferentes procedimientos de estimación considerados ( $\lambda = -1/2, 0, 2/3, 1, 2, 3$ ). Los errores estándar de las estimaciones se muestran en la Tabla 3.7.

Tabla 3.6. Estimadores de los parámetros

<b>Parámetros</b>	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
<i>Intercepto</i>	-0.2243	-0.2162	-0.2132	-0.2125	-0.2101	-0.1914
<i>centro_1</i>	-0.6473	-0.6489	-0.6500	-0.6508	-0.6576	-0.7014
<i>sexo_F</i>	0.1265	0.1311	0.1325	0.1327	0.1322	0.1259
<i>tratamiento_A</i>	1.2611	1.2624	1.2622	1.2616	1.2550	1.2100
<i>tratamiento_P</i>	-0.0182	-0.0184	-0.0185	-0.0185	-0.0184	-0.0179
<i>edad</i>	1.8653	1.8566	1.8542	1.8541	1.8573	1.8860

Tabla 3.7. Errores estándar de las estimaciones

<b>Parámetros</b>	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
<i>Intercepto</i>	0.5779	0.5776	0.5775	0.5774	0.5767	0.5731
<i>centro_A</i>	0.3524	0.3527	0.3527	0.3527	0.3523	0.3510
<i>sexo_F</i>	0.4401	0.4401	0.4402	0.4402	0.4405	0.4441
<i>tratamiento_A</i>	0.3459	0.3462	0.3463	0.3463	0.3461	0.3469
<i>tratamiento_P</i>	0.0129	0.0129	0.0129	0.0129	0.0129	0.0129
<i>edad</i>	0.3450	0.3454	0.3455	0.3455	0.3452	0.3452

Las estimaciones obtenidas con los diferentes procedimientos son análogas pero para la mayoría de los parámetros, los errores estándar son menores para la mayoría de los parámetros cuando se consideran otros procedimientos tales como los correspondientes a  $\lambda = -1/2, 0, 2, 3$ , que para el procedimiento EEG basado en el residuo de Pearson ( $\lambda = 1$ ).

### 3.4.3. Ejemplo 3: Uso de las EEG para Datos de Conteo

Algunas veces, los datos categóricos vienen en la forma de datos de conteo. Por ejemplo, podemos grabar el número de episodios de dolor agudo en un intervalo de tiempo en una prueba clínica para la evaluación de tratamientos. Otros ejemplos pueden ser: el número de reclamaciones de seguros registradas durante el año o el número de visitas médicas no previstas hechas durante un estudio de un nuevo protocolo para la medicación del asma. Con frecuencia, la regresión de Poisson es la estrategia apropiada para analizar tales datos.

Ya que la regresión de Poisson es una aplicación del modelo lineal general con la distribución de Poisson y la función de enlace log, podemos estimar modelos para cluster o datos repetidos con el método EEG. En este ejemplo, los investigadores evalúan un nuevo medicamento en el tratamiento de la osteoporosis en mujeres que han pasado la menopausia. En este estudio se utilizó el método doble-ciego, a un grupo de mujeres se les suministró el tratamiento y a otro grupo de mujeres se les suministró el placebo. A ambos grupos se les proporcionó calcio suplementario, dándoles consejos nutricionales, y animándolas a realizar ejercicio físico a través de un programa de ejercicios.

El estudio se realizó durante tres años, y el número de fracturas ocurridas fue anotado cada uno de estos años. La duración del correspondiente periodo de riesgo es de 12 meses. Sin embargo, hubo algunos abandonos en el tercer año, y estos periodos de riesgo fueron de 6 meses. La variable de compensación *lmonths*, es el logaritmo de los meses de riesgo

```
data fracture;
    input ID edad centro $ tratamiento $ año1 año2 año3 @@;
    total=año1+año2+año3;
    lmonths=log(12);
    datalines;
1 56 A p 0 0 0 2 71 A p 1 0 0 3 60 A p 0 0 1 4 71 A p 0 1 0
5 78 A p 0 0 0 6 67 A p 0 0 0 7 49 A p 0 0 0
9 75 A p 1 0 0 8 68 A p 0 0 0 11 82 A p 0 0 0
13 56 A p 0 0 0 12 71 A p 0 0 0 15 66 A p 1 0 0
```

17 78 A p 0 0 0 16 63 A p 0 2 0 19 61 A p 0 0 0  
 21 75 A p 1 0 0 20 68 A p 0 0 0 23 63 A p 1 1 1  
 25 54 A p 0 0 0 24 65 A p 0 0 0 27 71 A p 0 0 0  
 29 56 A p 0 0 0 28 64 A p 0 0 0 31 78 A p 0 0 2  
 33 76 A p 0 0 0 32 61 A p 0 0 0 35 76 A p 0 0 0  
 37 74 A p 0 0 0 36 56 A p 0 0 0 39 62 A p 0 0 0  
 41 56 A p 0 0 0 40 72 A p 0 0 1 43 76 A p 0 0 0  
 45 75 A p 0 0 0 44 77 A p 2 2 0 47 78 A p 0 0 0  
 49 71 A p 0 0 0 48 68 A p 0 0 0 51 74 A p 0 0 0  
 53 69 A p 0 0 0 52 78 A p 1 0 0 55 81 A p 2 0 1  
 57 68 A p 0 0 0 56 77 A p 0 0 0 59 77 A p 0 0 0  
 61 75 A p 0 0 0 60 83 A p 0 0 0 63 72 A p 0 0 0 64 88 A p 0 0 0  
 65 69 A p 0 0 0 66 55 A p 0 0 0 67 76 A p 0 0 0 68 55 A p 0 0 0  
 69 63 A t 0 0 2 70 52 A t 0 0 0 71 56 A t 0 0 0 72 52 A t 0 0 0  
 73 74 A t 0 0 0 74 61 A t 0 0 0 75 69 A t 0 0 0 76 61 A t 0 0 0  
 77 84 A t 0 0 0 78 76 A t 0 1 0 79 59 A t 0 0 1 80 76 A t 0 0 0  
 81 66 A t 0 0 1 82 78 A t 0 0 1 83 77 A t 0 0 0 84 75 A t 1 0 0  
 85 75 A t 0 0 0 86 62 A t 0 0 0 87 67 A t 0 0 0 88 62 A t 0 0 0  
 89 71 A t 0 0 0 90 63 A t 0 0 0 92 68 A t 0 0 0  
 93 69 A t 0 0 0 94 61 A t 0 0 0 96 61 A t 0 0 0  
 97 67 A t 0 0 0 98 77 A t 0 0 0 91 70 A t 0 0 1 102 81 A t 0 0 0  
 95 49 A t 0 0 0 106 55 A t 0 0 0  
 99 63 A t 2 1 0 100 52 A t 0 0 0 101 48 A t 0 0 0  
 103 71 A t 0 0 0 104 61 A t 0 0 0 105 74 A t 0 0 0  
 107 67 A t 0 0 0 108 56 A t 0 0 0 109 54 A t 0 0 0  
 111 56 A t 0 0 0 112 77 A t 1 0 0 113 65 A t 0 0 0  
 115 66 A t 0 0 0 116 71 A t 0 0 0 117 71 A t 0 0 0 128 71 A t 0 0 0  
 119 86 A t 1 0 0 120 81 A t 0 0 0 121 64 A t 0 0 0 132 76 A t 0 0 0  
 123 71 A t 0 0 0 124 76 A t 0 0 0 125 66 A t 0 0 0 136 76 A t 0 0 0  
 1 68 B p 0 0 0 2 63 B p 0 0 0 3 66 B p 0 0 0 4 63 B p 0 0 0  
 5 70 B p 0 1 0 6 62 B p 0 0 0 7 54 B p 1 0 0 8 66 B p 0 0 0

```

9   71 B p 0 0 0 10 76 B p 0 0 0 11 72 B p 0 0 1 12 65 B p 0 1 0
13  55 B p 0 1 0 14 59 B p 0 0 2 15 61 B p 1 0 0 16 56 B p 0 1 0
17  54 B p 0 0 0 18 68 B p 0 0 0 19 68 B p 0 0 0 20 81 B p 0 0 0
21  81 B p 1 0 0 22 61 B p 2 0 1 23 72 B p 1 0 0 24 67 B p 0 0 0
25  56 B p 0 0 0 26 66 B p 0 0 0 27 71 B p 0 1 0 28 75 B p 0 1 0
29  76 B p 0 0 0 30 73 B p 2 0 0 31 56 B p 0 0 0 32 89 B p 0 0 0
33  56 B p 0 0 0 34 78 B p 0 0 0 35 55 B p 0 0 0 36 73 B p 0 0 1
37  71 B p 0 0 0 38 56 B p 0 0 0 39 69 B p 0 0 0 40 77 B p 0 0 0
41  89 B p 0 0 0 42 63 B p 0 0 0 43 67 B p 0 0 0 44 73 B p 0 0 0
45  60 B p 0 0 0 46 67 B p 0 0 0 47 56 B p 0 0 0 48 78 B p 0 0 0
49  73 B t 1 0 0 50 76 B t 0 0 0 51 61 B t 0 0 0 52 81 B t 0 0 0
53  55 B t 0 0 0 54 82 B t 0 0 0 55 78 B t 0 0 0 56 60 B t 0 0 0
57  56 B t 0 0 0 58 83 B t 0 0 0 59 55 B t 0 0 0 60 60 B t 0 0 0
61  80 B t 0 0 0 62 78 B t 0 0 0 63 67 B t 0 0 0 64 67 B t 0 0 0
65  56 B t 0 0 0 66 72 B t 0 0 0 67 71 B t 0 0 0 68 83 B t 0 0 0
69  66 B t 0 0 0 70 71 B t 0 0 1 71 78 B t 1 0 2 72 61 B t 0 0 0
73  56 B t 0 0 0 74 61 B t 0 0 0 75 55 B t 0 0 0 76 69 B t 1 1 0
77  71 B t 0 0 0 78 76 B t 0 0 0 79 56 B t 0 0 0 80 75 B t 0 0 0
81  89 B t 0 0 0 82 77 B t 0 0 0 83 77 B t 1 0 0 84 73 B t 0 0 0
85  60 B t 0 0 0 86 61 B t 0 0 0 87 79 B t 0 0 0 88 71 B t 0 0 0
89  61 B t 0 0 0 90 79 B t 0 0 0 91 87 B t 1 0 0 92 55 B t 0 0 0
93  55 B t 0 0 0 94 79 B t 0 0 0 95 66 B t 0 0 0 96 49 B t 0 0 0
97  56 B t 0 0 0 98 64 B t 0 0 0 99 88 B t 0 0 0 100 62 B t 1 0 0
101 80 B t 0 0 1 102 65 B t 0 0 0 103 57 B t 0 0 1 104 85 B t 0 0 0
;

```

Para el análisis de Poisson en EEG especificamos la función enlace, LINK=LOG, y también especificamos la distribución, DIST=POISSON. La variable respuesta es *fracturas* y las variables en el modelo *centro*, *tratamiento*, *edad*, *año*, la interacción *tratamiento*×*centro* y la interacción *tratamiento*×*edad*. Además, en la regresión de Poisson normalmente se especifica una variable compensatoria. En esta situación, está la



variable *lmonths* que es el logaritmo de la duración del riesgo en cada año. La compensación se especifica con la opción `OFFSET=LMONTHS` en la sentencia `MODEL`. Ya que no hay un único identificador del individuo, usamos el cruce de *Id* y *centro* con la opción *subject* para crear valores únicos que determinen la unidad experimental.

Para estudiar este ejemplo primero lo implementamos en SAS utilizando el procedimiento estándar de SAS PROC GENMOD para calcular el valor inicial de los parámetros y después lo implementamos con los diferentes procedimientos alternativos a las EEG.

Las Tablas 3.8 y 3.9 muestran las estimaciones de los parámetros y el error estándar obtenido utilizando los nuevos procedimientos, respectivamente. Obsérvese que al considerar la familia de residuos basada en la familia de Cressie y Read se tiene un nuevo procedimiento para cada valor de  $\lambda$ , así como que  $\lambda = 1$  se corresponde con el procedimiento EEG.

Tabla 3.8. Estimadores de los parámetros

<b>Parámetros</b>	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
<i>Intercepto</i>	-6.7225	-6.6319	-6.6232	-6.6462	-8.500	-5.9511
<i>centro_A</i>	0.0402	0.0413	0.0414	0.0411	0.0173	0.0501
<i>tratamiento_P</i>	0.5549	0.5746	0.5765	0.5715	0.1687	0.7226
<i>edad</i>	0.0239	0.0225	0.0223	0.0227	0.0529	0.0114
<i>año_1</i>	0.2763	0.2763	0.2763	0.2763	0.2763	0.2763
<i>año_2</i>	-0.3830	-0.3830	-0.3830	-0.3830	-0.3830	-0.3830

Tabla 3.9. Errores estándar de las estimaciones

Parámetros	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
<i>Intercepto</i>	1.1360	1.1264	1.1257	1.1278	1.7110	1.1398
<i>centro_A</i>	0.3025	0.3010	0.3009	0.3011	0.4740	0.3195
<i>tratamiento_P</i>	0.3113	0.3094	0.3092	0.3096	0.4642	0.3191
<i>edad</i>	0.0150	0.0149	0.0149	0.0149	0.0236	0.0152
<i>año_1</i>	0.2992	0.2992	0.2992	0.2992	0.2992	0.2992
<i>año_2</i>	0.3814	0.2992	0.3814	0.3814	0.3814	0.3814

En este ejemplo los menores errores estándar para la mayoría de los parámetros se obtienen para  $\lambda = 2/3$  y  $\lambda = 0$ .

#### Conclusiones:

El esfuerzo computacional de todos los procedimientos incluido el clásico de Liang y Zeger es el mismo. En principio, las estimaciones no varían demasiado en estos ejemplos.

No obstante apuntan como los mejores procedimientos en la estimación de los parámetros los correspondientes a  $\lambda = -0.5$  y  $\lambda = 0$ .

En la Sección 3.5 se lleva a cabo un estudio de simulación necesario para conocer el verdadero alcance de los nuevos procedimientos.

### 3.5. Estudio de simulación

En la sección anterior se ha visto el comportamiento de los nuevos procedimientos alternativos a las EEG para modelizar tres conjuntos diferentes de datos reales de distinta índole utilizando como criterio la comparación de los errores estándar de los parámetros. Sin embargo, nos parece obligado llevar a cabo un estudio de simulación amplio para comparar estos nuevos procedimientos evaluando la consistencia de  $\beta$  así como la de los estimadores de su varianza.

El estudio se lleva a cabo bajo dos modelos distintos, uno de variables respuesta binaria y otro de variables respuesta Poisson que fueron anteriormente considerados en

el trabajo de Park et al. [84]. A continuación se detallan los pasos seguidos en el estudio de simulación:

**Primer paso:** A partir del modelo marginal se calculan las probabilidades de las variables respuesta.

**Segundo paso:** Generaremos variables correlacionadas con respuesta binaria o Poisson.

**Tercer paso:** Se calcula la estimación inicial del parámetro sin tener en cuenta la correlación entre las variables.

**Cuarto Paso:** Implementaremos los nuevos procedimientos alternativos a las EEG en SAS, basados en los residuos de  $\phi$ -divergencia definidos en (3.5) y en (3.6) para variables binarias y Poisson, respectivamente con  $\phi_{(\lambda)}(x)$  definido en (3.7) y con  $\lambda = -1/2, 0, 2/3, 1, 2$ . Se obtienen los estimadores de los parámetros del modelo simulado.

**Quinto Paso:** Repetimos 2000 veces los Pasos 2  $\rightarrow$  4 para muestras de 25, 50 y 100 individuos, con tres estructuras diferentes de correlación. Para comparar las estimaciones de los diferentes procedimientos se calcularán: el valor medio de todas ellas, el error cuadrático medio y las probabilidades de cubrimiento al 95 %.

### 3.5.1. Modelo para variables binarias correlacionadas

Consideramos medidas repetidas tomadas en tres tiempos diferentes ( $t=3$ ) obtenidas de dos grupos donde las variables respuesta son variables binarias correlacionadas. Se supone que el modelo marginal viene dado por:

$$\text{logit}(\mu_{ij}) = \log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \beta_0 + \beta_1 x_i + \beta_2 j + \beta_3 x_i j \quad (3.8)$$

donde

$$x_i = \begin{cases} 0 & \text{grupo 1} \\ 1 & \text{grupo 2} \end{cases} \quad y \quad j = 1, 2, 3$$

quedando

$$\left\{ \begin{array}{ll} \text{logit}(\mu_{ij}) = \beta_0 + \beta_2 j & \text{grupo 1} \\ \text{logit}(\mu_{ij}) = \beta_0 + \beta_1 + (\beta_2 + \beta_3)j & \text{grupo 2} \end{array} \right.$$

siendo

$\beta_0 \Rightarrow$  el intercepto del grupo 1.

$\beta_1 \Rightarrow$  el incremento del intercepto del grupo 2.

$\beta_2 \Rightarrow$  el efecto del tiempo en el grupo 1.

$\beta_3 \Rightarrow$  el incremento del efecto del tiempo en el grupo 2.

Los verdaderos valores de los parámetros son elegidos como

$$\beta_T = (\beta_0, \beta_1, \beta_2, \beta_3) = (0.1, 0.2, 0.2, 0.1).$$

Para simplificar asumimos que no hay sobredispersión, o lo que es lo mismo, que  $\gamma = 1$ . Las estructuras de  $\mathbf{R}$  que vamos a considerar para estudiar los efectos de las diferentes estructuras de correlación (ver Tabla 3.2, en la sección anterior) en los dos grupos son:

1. desestructurada con coeficientes de correlación 0.5, 0.25 y 0.5 en ambos grupos.
2. estructura intercambiable en el grupo 1 con  $\alpha = 0.5$  y estructura de correlación desestructurada en el grupo 2 con coeficientes de correlación 0.5, 0.25 y 0.5.
3. estructura intercambiable en el grupo 1 con  $\alpha = 0.1$  y estructura de correlación desestructurada en el grupo 2 con coeficientes de correlación 0.5, 0.25 y 0.5.

A partir de aquí, una vez que hemos calculado la probabilidad asociada para cada grupo y fijada una estructura de correlación, vamos a utilizar el método propuesto por Park et al. [83] para generar variables correlacionadas con respuesta binaria que llamamos a continuación.

Se propone un algoritmo para generar un vector aleatorio  $(Z_1, \dots, Z_k)^T$  de variables binarias tales que  $E[Z_i] = p_i$ ,  $q_i = 1 - p_i$  y  $\text{corr}(Z_i, Z_j) = \rho_{ij} \geq 0$ ,  $i \neq j$ . Sea  $\mathbf{X}(\alpha)$  una variable aleatoria Poisson teniendo media  $\alpha \geq 0$ . Por convenio  $X(0) = 0$ . De aquí

en adelante asumimos que las  $X(\cdot)$  son independientes unas de otras si ellas aparecen con distintos subíndices.

**Algoritmo:**

**Paso 0:** Calculamos

$$\alpha_{ij} = \log \left[ 1 + \rho_{ij} \{q_i p_i^{-1} q_j p_j^{-1}\}^{1/2} \right]$$

para  $1 \leq i, j \leq k$ .

Sea  $l = 0$  y  $\alpha_{ij}^1 = \alpha_{ij}$ ,  $1 \leq i, j \leq k$ .

**Paso 1:** Sea  $l = l + 1$ .

Determinamos  $\beta_l$  y  $(r, s)$ .

Definimos

$$T_l = \{\alpha_{ij}^l : \alpha_{ij}^l > 0, 1 \leq i, j \leq k\}.$$

Sea  $\beta_l = \alpha_{rs}^l$  el menor elemento del conjunto  $T_l$ . Si  $\alpha_{rr} = 0$  ó  $\alpha_{ss} = 0$ , entonces parar.

Calculamos  $S_l$  :

Sea  $S_l^0 = \{r, s\}$

Para  $i = 1, 2, \dots, k$ ; sea

$$\begin{aligned} S_l^i &= S_l^{i-1} \cup \{i\} && \text{si } \alpha_{ij}^l > 0 \text{ para todo } j \in S_l^{i-1} \\ &= S_l^{i-1} && \text{en caso contrario} \end{aligned}$$

Sea  $S_l = S_l^k$

**Paso 2:** Actualizamos los  $\alpha_{ij}$  :

$$\begin{aligned} \alpha_{ij}^{l+1} &= \alpha_{ij}^l - \beta_l, && \forall i \in S_l \text{ y } j \in S_l \\ &= \alpha_{ij}^l, && \forall i \notin S_l \text{ o } j \notin S_l. \end{aligned}$$

Si todos los  $\alpha_{ij}^{l+1} = 0$  para  $1 \leq i, j \leq k$ , entonces vamos al Paso 3. En caso contrario vamos al Paso 1.

**Paso 3:** Sea  $\tau = l$ .

$$Y_i = \sum_{l=1}^{\tau} X_l (\beta_l) I_{S_l} (i); \text{ para } i = 1, \dots, k$$

y el conjunto

$$Z_i = I_{\{0\}} (Y_i)$$

$$\text{donde } I_A \text{ es la función indicatriz de un conjunto } A \text{ tal que } I_A (y) = \begin{cases} 1 & y \in A \\ 0 & y \notin A \end{cases}$$

Nótese que en el Paso 1 el mínimo  $\beta_l$  es elegido de tal manera que asegura que todas las variables aleatorias Poisson tienen media no negativa. Es posible que  $\alpha_{rs}$  y  $S_l$  en el Paso 1 puedan no ser determinadas unívocamente. En este caso podemos elegir  $\alpha_{rs}$  y  $S_l$  arbitrariamente.

Calculamos el vector aleatorio binario  $\mathbf{Z} = (Z_1, Z_2, Z_3)$  a partir del algoritmo anterior generando tantas variables Poisson independientes de parámetro  $\beta_l$  como sean necesarias.

La comparación de los diferentes procedimientos se lleva a cabo utilizando las siguientes medidas:

El **valor promedio** de los parámetros estimados como

$$\sum_{k=1}^{2000} \frac{\hat{\beta}_{k,i}}{2000}$$

donde  $\hat{\beta}_{k,i}$  es el estimador del parámetro  $\beta_i$ ,  $i = 0, \dots, 3$  de la replica  $k$  -ésima.

El **error cuadrático medio** ( $ECM_i$ ) de  $\hat{\beta}_{k,i}$

$$ECM_i = \sqrt{\sum_{k=1}^{2000} \frac{(\hat{\beta}_{k,i} - \beta_T)^2}{2000}}$$

donde  $\beta_T$  es el verdadero valor del parámetro  $\beta$ .

La **probabilidad de cubrimiento** al 95 %, se calcula como la proporción de las réplicas en las que el Intervalo de Confianza al 95 % (IC) contiene el verdadero valor del parámetro. Para cada réplica calculamos el IC de  $\beta_i$  de la forma

$$\hat{\beta}_{k,i} \pm 1,96s_i$$

donde  $s_i$  el error estándar estimado de  $\hat{\beta}_{k,i}$ .

La Tabla 3.10 muestra la salidas de la simulación en el caso en el que la estructura de la matriz de correlación de trabajo sea desestructurada con coeficientes de correlación 0.5, 0.25 y 0.5 en ambos grupos. Para los tres criterios de comparación, en la mayoría de los casos, el mejor de los procedimientos de estimación se corresponde con  $\lambda = -1/2$ . Destacar que únicamente el procedimiento clásico EEG ( $\lambda = 1$ ) es el mejor en 2 de 12 casos para el primer criterio de comparación, nunca es el mejor para el criterio de ECM y sólo es mejor en 1 de 12 casos para el criterio de la probabilidad de cubrimiento.

La Tabla 3.11 muestra la salidas de la simulación en el caso en el que la estructura de la matriz de correlación de trabajo sea intercambiable en el grupo 1 con  $\alpha = 0.5$  y estructura de correlación desestructurada en el grupo 2 con coeficientes de correlación 0.5, 0.25 y 0.5. De nuevo el procedimiento de estimación que se corresponde con  $\lambda = -1/2$  es el mejor o el segundo mejor en la mayoría de los casos. El procedimiento clásico ( $\lambda = 1$ ) es a lo sumo mejor en 3 casos de los 12 para dos de los tres criterios de comparación.

La Tabla 3.12 muestra la salidas de la simulación en el caso en el que la estructura de la matriz de correlación de trabajo sea intercambiable en el grupo 1 con  $\alpha = 0.1$  y estructura de correlación desestructurada en el grupo 2 con coeficientes de correlación 0.5, 0.25 y 0.5. La conclusión coincide con las tablas anteriores, aunque no siempre es el mejor procedimiento de acuerdo a los tres criterios de comparación, el correspondiente a  $\lambda = -1/2$  es el que en más casos es mejor que el resto de procedimientos. El procedimiento clásico ( $\lambda = 1$ ) únicamente es mejor en uno y 2 casos de 12 para el primer y segundo criterio. Para el tercer criterio nunca es mejor.

Tabla 3.10. Resultados simulación para respuestas binarias con matriz de correlación de trabajo desestructurada en ambos grupos

$n$			$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
100	$\beta_0$	<i>Media</i>	0.1016	0.1010	0.1002	0.1000	0.1004
		<i>ECM</i>	0.2889	0.2890	0.2893	0.2891	0.2895
		<i>Pr ob Cub</i>	0.9540	0.9530	0.9525	0.9510	0.9520
	$\beta_1$	<i>Media</i>	0.2075	0.2062	0.2066	0.2066	0.2071
		<i>ECM</i>	0.4315	0.4319	0.4324	0.4322	0.4330
		<i>Pr ob Cub</i>	0.9500	0.9500	0.9505	0.9495	0.9510
	$\beta_2$	<i>Media</i>	0.2029	0.2032	0.2034	0.2033	0.2035
		<i>ECM</i>	0.1247	0.1247	0.1247	0.1248	0.1246
		<i>Pr ob Cub</i>	0.9545	0.9550	0.9550	0.9540	0.9550
	$\beta_3$	<i>Media</i>	0.0968	0.0974	0.0974	0.0970	0.0973
		<i>ECM</i>	0.1927	0.1927	0.1927	0.1927	0.1928
		<i>Pr ob Cub</i>	0.9425	0.9415	0.9415	0.9405	0.9420
50	$\beta_0$	<i>Media</i>	0.0984	0.0966	0.0963	0.0962	0.0965
		<i>ECM</i>	0.4098	0.4113	0.4117	0.4118	0.4125
		<i>Pr ob Cub</i>	0.9565	0.9565	0.9565	0.9570	0.9565
	$\beta_1$	<i>Media</i>	0.2139	0.2142	0.2136	0.2143	0.2143
		<i>ECM</i>	0.6049	0.6077	0.6082	0.6087	0.6096
		<i>Pr ob Cub</i>	0.9555	0.9575	0.9580	0.9585	0.9585
	$\beta_2$	<i>Media</i>	0.2030	0.2039	0.2040	0.2041	0.2042
		<i>ECM</i>	0.1818	0.1822	0.1823	0.1823	0.1824
		<i>Pr ob Cub</i>	0.9475	0.9480	0.9460	0.9465	0.9465
	$\beta_3$	<i>Media</i>	0.1027	0.1027	0.1027	0.1027	0.1025
		<i>ECM</i>	0.2664	0.2673	0.2674	0.2674	0.2673
		<i>Pr ob Cub</i>	0.9510	0.9500	0.9460	0.9495	0.9485
25	$\beta_0$	<i>Media</i>	0.0958	0.0926	0.0920	0.0920	0.0928
		<i>ECM</i>	0.5918	0.5938	0.5946	0.5948	0.5952
		<i>Pr ob Cub</i>	0.9615	0.9485	0.9595	0.9595	0.9600
	$\beta_1$	<i>Media</i>	0.2108	0.2082	0.2076	0.2076	0.2077
		<i>ECM</i>	0.8724	0.8762	0.8775	0.8777	0.8782
		<i>Pr ob Cub</i>	0.9490	0.9485	0.9480	0.948	0.9485
	$\beta_2$	<i>Media</i>	0.2098	0.2112	0.2115	0.2115	0.2113
		<i>ECM</i>	0.2684	0.2691	0.2694	0.2694	0.2692
		<i>Pr ob Cub</i>	0.9385	0.9405	0.9400	0.9400	0.9400
	$\beta_3$	<i>Media</i>	0.1106	0.1119	0.1123	0.1123	0.1124
		<i>ECM</i>	0.3901	0.3918	0.3923	0.3923	0.3922
		<i>Pr ob Cub</i>	0.9495	0.9490	0.9490	0.9490	0.9475



Tabla 3.11. Resultados simulación para respuestas binarias con matriz de correlación de trabajo intercambiable en el grupo 1 con  $\alpha=0.5$  y estructura de correlación desestructurada en el grupo 2 con coeficientes de correlación 0.5, 0.25 y 0.5.

$n$			$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
100	$\beta_0$	<i>Media</i>	0.1040	0.1040	0.1039	0.1040	0.1036
		<i>ECM</i>	0.2716	0.2712	0.2712	0.2711	0.2706
		<i>Pr ob Cub</i>	0.9410	0.9395	0.9405	0.9410	0.9410
	$\beta_1$	<i>Media</i>	0.2046	0.2034	0.2037	0.2031	0.2032
		<i>ECM</i>	0.4146	0.4146	0.4150	0.4147	0.4149
		<i>Pr ob Cub</i>	0.9520	0.9535	0.9535	0.9540	0.9530
	$\beta_2$	<i>Media</i>	0.1999	0.2002	0.2003	0.2003	0.2003
		<i>ECM</i>	0.1058	0.1057	0.1058	0.1057	0.1058
		<i>Pr ob Cub</i>	0.9485	0.9505	0.9490	0.9490	0.9485
	$\beta_3$	<i>Media</i>	0.1002	0.1006	0.1005	0.1006	0.1006
		<i>ECM</i>	0.1782	0.1783	0.1784	0.1783	0.1783
		<i>Pr ob Cub</i>	0.9410	0.9415	0.9425	0.9425	0.9425
50	$\beta_0$	<i>Media</i>	0.0928	0.0917	0.0917	0.0918	0.0916
		<i>ECM</i>	0.3674	0.3673	0.3675	0.3675	0.3678
		<i>Pr ob Cub</i>	0.9495	0.9515	0.9530	0.9520	0.9505
	$\beta_1$	<i>Media</i>	0.2191	0.2188	0.2186	0.2182	0.2192
		<i>ECM</i>	0.5804	0.5814	0.5818	0.5818	0.5821
		<i>Pr ob Cub</i>	0.9535	0.9540	0.9550	0.9545	0.9540
	$\beta_2$	<i>Media</i>	0.2050	0.2058	0.2060	0.2059	0.2058
		<i>ECM</i>	0.1461	0.1461	0.1460	0.1461	0.1462
		<i>Pr ob Cub</i>	0.9495	0.9500	0.9505	0.9500	0.9490
	$\beta_3$	<i>Media</i>	0.1010	0.1007	0.1008	0.1009	0.1005
		<i>ECM</i>	0.2476	0.2479	0.2480	0.2478	0.2476
		<i>Pr ob Cub</i>	0.9505	0.9530	0.9535	0.9530	0.9520
25	$\beta_0$	<i>Media</i>	0.0908	0.0904	0.0901	0.0903	0.0909
		<i>ECM</i>	0.5371	0.5375	0.5380	0.5381	0.5395
		<i>Pr ob Cub</i>	0.9575	0.9580	0.9580	0.9580	0.9580
	$\beta_1$	<i>Media</i>	0.2136	0.2096	0.2092	0.2088	0.2089
		<i>ECM</i>	0.8375	0.8405	0.8415	0.8419	0.8433
		<i>Pr ob Cub</i>	0.9535	0.9530	0.9540	0.9550	0.9530
	$\beta_2$	<i>Media</i>	0.2068	0.2081	0.2083	0.2083	0.2083
		<i>ECM</i>	0.2192	0.2192	0.2193	0.2193	0.2191
		<i>Pr ob Cub</i>	0.9415	0.9450	0.9435	0.9435	0.9445
	$\beta_3$	<i>Media</i>	0.1144	0.1154	0.1155	0.1157	0.1155
		<i>ECM</i>	0.3562	0.3574	0.3577	0.3578	0.3572
		<i>Pr ob Cub</i>	0.9520	0.9520	0.9520	0.9515	0.9490

Tabla 3.12. Resultados simulación para respuestas binarias con matriz de correlación de trabajo intercambiable en el grupo 1 con  $\alpha=0.1$  y estructura de correlación desestructurada en el grupo 2 con coeficientes de correlación 0.5, 0.25 y 0.5.

$n$			$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
100	$\beta_0$	<i>Media</i>	0.1020	0.1024	0.1023	0.1023	0.1022
		<i>ECM</i>	0.3007	0.3003	0.3003	0.3003	0.3003
		<i>Pr ob Cub</i>	0.9530	0.9520	0.9525	0.9530	0.9525
	$\beta_1$	<i>Media</i>	0.2065	0.2049	0.2045	0.2048	0.2050
		<i>ECM</i>	0.4362	0.4363	0.4363	0.4365	0.4366
		<i>Pr ob Cub</i>	0.9510	0.9515	0.9515	0.9520	0.9515
	$\beta_2$	<i>Media</i>	0.1998	0.1997	0.1998	0.1999	0.2000
		<i>ECM</i>	0.1388	0.1388	0.1389	0.1388	0.1388
		<i>Pr ob Cub</i>	0.9535	0.9520	0.9520	0.9525	0.9545
	$\beta_3$	<i>Media</i>	0.1004	0.1009	0.1010	0.1009	0.1009
		<i>ECM</i>	0.2010	0.2010	0.2011	0.2011	0.2011
		<i>Pr ob Cub</i>	0.9395	0.9375	0.9385	0.9390	0.9390
50	$\beta_0$	<i>Media</i>	0.0925	0.0923	0.0925	0.0922	0.0920
		<i>ECM</i>	0.4281	0.4275	0.4276	0.4276	0.4279
		<i>Pr ob Cub</i>	0.9500	0.9495	0.9495	0.9495	0.9480
	$\beta_1$	<i>Media</i>	0.2181	0.2177	0.2173	0.2177	0.2186
		<i>ECM</i>	0.6198	0.6202	0.6206	0.6205	0.6203
		<i>Pr ob Cub</i>	0.9500	0.9495	0.9495	0.9495	0.9500
	$\beta_2$	<i>Media</i>	0.2075	0.2079	0.2080	0.2081	0.2081
		<i>ECM</i>	0.1989	0.1987	0.1988	0.1988	0.1990
		<i>Pr ob Cub</i>	0.9510	0.9530	0.9530	0.9525	0.9510
	$\beta_3$	<i>Media</i>	0.0990	0.0989	0.0990	0.0989	0.0983
		<i>ECM</i>	0.2841	0.2842	0.2843	0.2843	0.2838
		<i>Pr ob Cub</i>	0.9500	0.9495	0.9500	0.9500	0.9500
25	$\beta_0$	<i>Media</i>	0.0966	0.0973	0.0974	0.0973	0.0964
		<i>ECM</i>	0.6163	0.6170	0.6176	0.6178	0.6185
		<i>Pr ob Cub</i>	0.9530	0.9560	0.9560	0.9570	0.9555
	$\beta_1$	<i>Media</i>	0.2076	0.2030	0.2020	0.2020	0.2038
		<i>ECM</i>	0.8962	0.8987	0.8998	0.9001	0.9008
		<i>Pr ob Cub</i>	0.9485	0.9490	0.9485	0.9480	0.9480
	$\beta_2$	<i>Media</i>	0.2021	0.2024	0.2025	0.2026	0.2033
		<i>ECM</i>	0.2791	0.2794	0.2797	0.2798	0.2804
		<i>Pr ob Cub</i>	0.9515	0.9515	0.9505	0.9500	0.9490
	$\beta_3$	<i>Media</i>	0.1193	0.1210	0.1213	0.1212	0.1204
		<i>ECM</i>	0.3999	0.4009	0.4014	0.4015	0.4016
		<i>Pr ob Cub</i>	0.9495	0.9515	0.9495	0.9500	0.9495

### Extensión a una nueva familia penalizada de residuos.

Tras hacer el estudio de simulación para el caso de variables aleatorias con respuesta binaria correlacionada en el que  $\lambda = -1/2$  se corresponde con el mejor procedimiento de estimación, nos planteamos la duda de si quizá valores más pequeños que  $\lambda = -1/2$  podrían proporcionar mejores procedimientos.

Debido a que la familia de residuos basada en la divergencia de Cressie y Read sólo es válida para valores  $\lambda > -1$ , decidimos utilizar una familia penalizada de esta familia de divergencias estudiada en otros contextos, ver Basu et al. [16] y Pardo et al. [80].

Se define la familia penalizada de residuos

$$c_{ij}^{\phi(\lambda),w} = \text{sig}(y_{ij} - \hat{\mu}_{ij}) \sqrt{\frac{2}{\phi''_{(\lambda),w}(1)}} \left\{ \hat{\mu}_{ij} \phi_{(\lambda),w} \left( \frac{y_{ij}}{\hat{\mu}_{ij}} \right) + (1 - \hat{\mu}_{ij}) \phi_{(\lambda),w} \left( \frac{1 - y_{ij}}{1 - \hat{\mu}_{ij}} \right) \right\}^{1/2}$$

donde

$$\phi_{(\lambda),w} = \begin{cases} w & \text{si } y_{ij} = 0 & \forall \lambda \\ x \log x - x + 1 & \text{si } y_{ij} \neq 0, & \lambda = 0 \\ -\log x + x - 1 & \text{si } y_{ij} \neq 0, & \lambda = -1 \\ \frac{1}{\lambda(1+\lambda)} \left( x^{\lambda+1} - x - (x-1)^\lambda \right) & \text{si } y_{ij} \neq 0, & \lambda \neq 0, \lambda \neq -1. \end{cases}$$

Para la simulación hemos tomado los valores de  $w = 1/2, 0, 1$  y  $\lambda = -2, -1, -1/2, 0, 2/3, 1, 2$ . Hemos considerado un tamaño muestral de  $n=25$  en ambos grupos.

Obsérvese que si  $w = 1$ , el residuo penalizado  $c_{ij}^{\phi(0),1}$  coincide con el residuo no penalizado  $c_{ij}^{\phi(0)}$ . Por otra parte, si  $w = 1/2$ , el residuo penalizado  $c_{ij}^{\phi(0),1/2}$  coincide con el residuo no penalizado  $c_{ij}^{\phi(1)}$ .

Las Tablas 3.13-3.15 muestran las salidas de la simulación para los tres estructuras de la matriz de correlación de trabajo que hemos considerado anteriormente.



Tabla 3.14. Resultados simulación para respuestas binarias con matriz de correlación de trabajo intercambiable en el grupo 1 con  $\alpha=0.5$  y estructura de correlación desestructurada en el grupo 2 con coeficientes de correlación 0.5, 0.25 y 0.5

$w$	$Par.$	$Estad.$	$\lambda = -2$	$\lambda = -1$	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
0	$\beta_0$	<i>Media</i>	0.0899	0.0900	0.0900	0.0900	0.0901	0.0897	0.0902
		<i>ECM</i>	0.5382	0.5382	0.5383	0.5383	0.5383	0.5389	0.5395
		<i>ProbCub</i>	0.9595	0.9590	0.9595	0.9595	0.9590	0.9580	0.9590
	$\beta_1$	<i>Media</i>	0.2065	0.2067	0.2068	0.2069	0.2071	0.2077	0.2084
		<i>ECM</i>	0.8433	0.8433	0.8432	0.8432	0.8433	0.8437	0.8441
		<i>ProbCub</i>	0.9540	0.9540	0.9535	0.9540	0.9545	0.9545	0.9530
	$\beta_2$	<i>Media</i>	0.2088	0.2088	0.2088	0.2088	0.2085	0.2088	0.2087
		<i>ECM</i>	0.2195	0.2195	0.2195	0.2195	0.2194	0.2194	0.2191
		<i>ProbCub</i>	0.9430	0.9430	0.9430	0.9430	0.9420	0.9425	0.9445
	$\beta_3$	<i>Media</i>	0.1163	0.1163	0.1162	0.1162	0.1161	0.1162	0.1158
		<i>ECM</i>	0.3588	0.3587	0.3587	0.3586	0.3585	0.3583	0.3577
		<i>ProbCub</i>	0.9520	0.9520	0.9520	0.9510	0.9505	0.9510	0.9495
1/2	$\beta_0$	<i>Media</i>	0.0901	0.0903	0.0898	0.0902	0.0903	0.0903	0.0905
		<i>ECM</i>	0.5378	0.5379	0.5377	0.5379	0.5380	0.5381	0.5395
		<i>ProbCub</i>	0.9595	0.9595	0.9585	0.9580	0.9580	0.9580	0.9580
	$\beta_1$	<i>Media</i>	0.2076	0.2077	0.2078	0.2081	0.2085	0.2087	0.2100
		<i>ECM</i>	0.8420	0.8420	0.8419	0.8419	0.8419	0.8419	0.8428
		<i>ProbCub</i>	0.9545	0.9545	0.9540	0.9545	0.9545	0.9550	0.9525
	$\beta_2$	<i>Media</i>	0.2085	0.2085	0.2084	0.2084	0.2083	0.2083	0.2082
		<i>ECM</i>	0.2193	0.2193	0.2193	0.2193	0.2193	0.2193	0.2191
		<i>ProbCub</i>	0.9445	0.9445	0.9435	0.9440	0.9435	0.9435	0.9440
	$\beta_3$	<i>Media</i>	0.1160	0.1158	0.1158	0.1158	0.1157	0.1157	0.1153
		<i>ECM</i>	0.3582	0.3581	0.3580	0.3580	0.3579	0.3578	0.3570
		<i>ProbCub</i>	0.9520	0.9520	0.9515	0.9520	0.9515	0.9515	0.9490
1	$\beta_0$	<i>Media</i>	0.0903	0.0904	0.0904	0.0904	0.0905	0.0905	0.0909
		<i>ECM</i>	0.5374	0.5374	0.5375	0.5375	0.5377	0.5378	0.5397
		<i>ProbCub</i>	0.9580	0.9580	0.9580	0.9580	0.9580	0.9580	0.9575
	$\beta_1$	<i>Media</i>	0.2089	0.2092	0.2094	0.2096	0.2100	0.2104	0.2120
		<i>ECM</i>	0.8406	0.8405	0.8405	0.8405	0.8405	0.8405	0.8416
		<i>ProbCub</i>	0.9540	0.9540	0.9535	0.9530	0.9530	0.9530	0.9505
	$\beta_2$	<i>Media</i>	0.2081	0.2080	0.2080	0.2080	0.2079	0.2078	0.2076
		<i>ECM</i>	0.2192	0.2192	0.2192	0.2192	0.2192	0.2192	0.2191
		<i>ProbCub</i>	0.9450	0.9450	0.9450	0.9450	0.9450	0.9450	0.9430
	$\beta_3$	<i>Media</i>	0.1156	0.1156	0.1155	0.1154	0.1153	0.1152	0.1148
		<i>ECM</i>	0.3576	0.3575	0.3574	0.3574	0.3572	0.3571	0.3564
		<i>ProbCub.</i>	0.9520	0.9520	0.9520	0.9520	0.9525	0.9525	0.9515



La Tabla 3.13 muestra la salidas de la simulación en el caso en el que la estructura de la matriz de correlación de trabajo sea desestructurada con coeficientes de correlación 0.5, 0.25 y 0.5 en ambos grupos. Utilizando el criterio de la media vemos que los valores que más se aproximan a los verdaderos valores de los parámetros los obtenemos, para cualquier valor de  $w$ , con  $\lambda = 2$ . El menor ECM lo obtenemos para cualquier valor de  $w$ , cuando  $\lambda = 2$ . Estudiando el criterio de la probabilidad de cubrimiento observamos que para  $w = 0$  los valores más próximos al 95 % los obtenemos cuando  $\lambda = -2, -1$  y  $-1/2$ , para  $w = 1/2$  cuando  $\lambda = 1$  y para  $w = 1$  cuando  $\lambda = 0, 2/3$ .

La Tabla 3.14 muestra la salidas de la simulación en el caso en el que la estructura de la matriz de correlación de trabajo sea intercambiable en el grupo 1 con  $\alpha = 0,5$  y estructura de correlación desestructurada en el grupo 2 con coeficientes de correlación 0.5, 0.25 y 0.5. Utilizando el criterio de la media vemos que los resultados que más se aproximan a los verdaderos valores de los parámetros los obtenemos, para cualquier valor de  $w$ , con  $\lambda = 2$ . El menor ECM lo obtenemos para cualquier valor de  $w$ , con  $\lambda = 2$ . Estudiando el criterio de la probabilidad de cubrimiento observamos que para cualquier valor de  $w$  los valores más próximos al 95 % los obtenemos cuando  $\lambda = 2$ .

La Tabla 3.15 muestra la salidas de la simulación en el caso en el que la estructura de la matriz de correlación de trabajo sea intercambiable en el grupo 1 con  $\alpha = 0,1$  y estructura de correlación desestructurada en el grupo 2 con coeficientes de correlación 0.5, 0.25 y 0.5. Utilizando el criterio de la media vemos que los resultados que más se aproximan a los verdaderos valores de los parámetros los obtenemos, para  $w = 0$  y 1 con  $\lambda = -2$  y para  $w = 1/2$  con  $\lambda = -2$  y 0. El menor ECM lo obtenemos cuando  $w = 0$  el mejor procedimiento sería  $\lambda = 2/3$ , cuando  $w = 1/2$  para  $\lambda = -2$ , y cuando  $w = 1$  para  $\lambda = -2, 0, 2/3, 1, 2$ . Estudiando el criterio de la probabilidad de cubrimiento observamos que para  $w = 0$  y 1 los valores más próximos al 95 % los obtenemos cuando  $\lambda = -1/2$  y para  $w = 1/2$  cuando  $\lambda = 1$ .

Por lo que podríamos concluir, que los mejores resultados, en la mayoría de los casos, para las matrices de correlación de trabajo con la estructura especificada y para cualquier valor de  $w$ , es el procedimiento obtenido cuando  $\lambda = 2$ .

Si se comparan la columna correspondiente a  $\lambda = 2$  de estas tablas obtenidas utilizando el residuo penalizado con las Tablas 3.10-3.12, se obtienen mejores resultados

utilizando  $c_{ij}^{\phi(-1/2)}$  que  $c_{ij}^{\phi(2),w}$ . Por tanto, nuestra duda se despejó, o lo que es lo mismo, no encontramos mejores procedimientos cuando  $\lambda < -1/2$ .

### 3.5.2. Modelo para variables Poisson correlacionadas

Consideramos medidas repetidas tomadas en tres tiempos diferentes (t=3) obtenidas de dos grupos donde las variables respuesta son variables Poisson correlacionadas. Se supone que el modelo marginal viene dado por:

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j + \beta_3 x_i j \quad (3.9)$$

donde

$$x_i = \begin{cases} 0 & \text{grupo 1} \\ 1 & \text{grupo 2} \end{cases} \quad y \quad j = 1, 2, 3$$

quedando

$$\begin{cases} \log(\mu_{ij}) = \beta_0 + \beta_2 j & \text{grupo 1} \\ \log(\mu_{ij}) = \beta_0 + \beta_1 + (\beta_2 + \beta_3)j & \text{grupo 2} \end{cases} \quad (3.10)$$

siendo

$\beta_0 \Rightarrow$  el intercepto del grupo 1.

$\beta_1 \Rightarrow$  el incremento del intercepto del grupo 2.

$\beta_2 \Rightarrow$  el efecto del tiempo en el grupo 1.

$\beta_3 \Rightarrow$  el incremento del efecto del tiempo en el grupo 2.

Los verdaderos valores de los parámetros son elegidos como

$$\beta_T = (\beta_0, \beta_1, \beta_2, \beta_3) = (0.1, 0.2, 0.2, 0.1).$$

Para simplificar asumimos que no hay sobredispersión, o lo que es lo mismo, que  $\gamma = 1$ . Las estructuras de  $\mathbf{R}$  que vamos a considerar para estudiar los efectos de las diferentes estructuras de correlación (ver Tabla 3.2, en la sección anterior) en los dos grupos son:

1. estructura intercambiable con  $\alpha = 0.5$ .



2. estructura AR-1 con  $\alpha=0.5$ .
3. desestructurada con coeficientes de correlación 0.5, 0.1, 0.3

A partir de aquí, una vez que hemos calculado la probabilidad asociada para cada grupo y fijada una estructura de correlación, vamos a utilizar el método propuesto por Park y Shin [85] para generar variables correlacionadas con respuesta Poisson que detallamos a continuación.

Generamos variables Poisson correlacionadas  $(\mathbf{Z}_1, \dots, \mathbf{Z}_k)'$  a partir de variables respuesta Poisson independientes.

**Paso 1:** Sea  $l = 1$

Calculamos  $\alpha_{ij}^l = \rho_{ij} (\mu_i \mu_j)^{1/2}$  para  $1 \leq i, j \leq k$ .

Construimos un grafo conectando los vértices  $i, j$  con  $\alpha_{ij} > 0$  donde cada vértice representa una de las  $k$  variables aleatorias. El valor de  $\alpha_{ij}$  se asigna a la arista  $(i, j)$  que une los vértices  $i, j$ .

**Paso 2:** A la arista con menor valor se denota por  $(r, s)$ .

Sea  $\beta_l = \alpha_{rs}^l$ .

**Paso 3:** Encontramos un conjunto  $S_l$  de vértices unidos que contengan los vértices  $r$  y  $s$ .

**Paso 4:** Actualizamos los valores de las aristas de la siguiente forma

$$\begin{aligned} \alpha_{ij}^{l+1} &= \alpha_{ij}^l - \beta_l, \quad \forall i \in S_l \text{ y } j \in S_l \\ &= \alpha_{ij}^l, \quad \forall i \notin S_l \text{ y } j \notin S_l \end{aligned}$$

Se quitan las aristas  $(i, j)$  con  $\alpha_{ij}^{l+1} = 0$ . Si todas las nuevas aristas tienen valor  $\leq 0$  entonces vamos al Paso 5.

Si no,  $l = l + 1$  y volvemos al paso 2.

**Paso 5:** Si algún vértice tiene valor  $\alpha_{ij}^{l+1} < 0$ , entonces el algoritmo falla y se para.

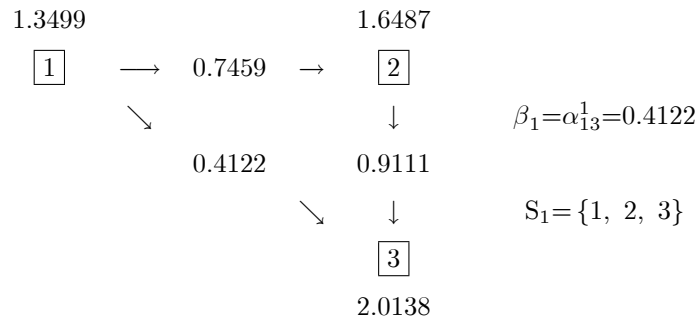
En caso contrario, se construye una matriz  $k \times \tau$ ,  $\mathbf{T} = [T_{il}]$  con

$$T_{il} = \begin{cases} 1, & \text{si } i \in S_l, \text{ para } i = 1, \dots, k, l = 1, \dots, \tau. \\ 0, & \text{si } i \notin S_l \end{cases}$$

Entonces,  $\mathbf{Z} = \mathbf{T} \times \mathbf{Y}$ , donde  $\mathbf{Y}$  está formado por variables independientes Poisson de parámetros  $\beta_l$ , son un conjunto de variables aleatorias Poisson correlacionadas.

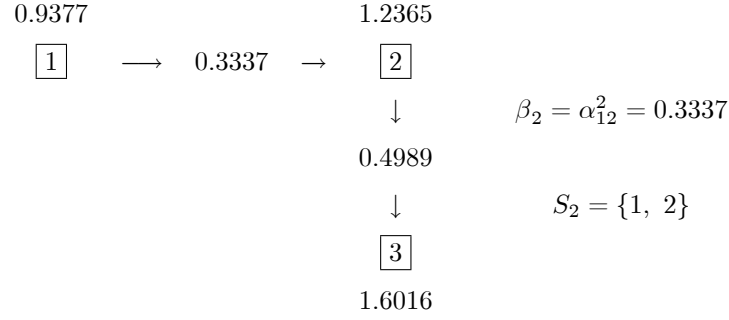
**Ejemplo de aplicación del algoritmo para el caso de matriz de correlación de trabajo AR-1.** Se considera el modelo sólo para el grupo 1. El vector de medias viene dado por  $\mu = (1.3499, 1.6487, 2.0138)$  y se consideran los coeficientes de correlación  $\rho = \{1 \ 0.5 \ 0.25, 0.5 \ 1 \ 0.5, 0.25 \ 0.5 \ 1\}$ . A continuación se ilustra gráficamente el algoritmo. Cada nodo del grafo representa una de las tres variables aleatorias.

**1=1** Calculamos los  $\alpha_{ij} = \rho_{ij} \sqrt{\mu_i \mu_j}$  iniciales



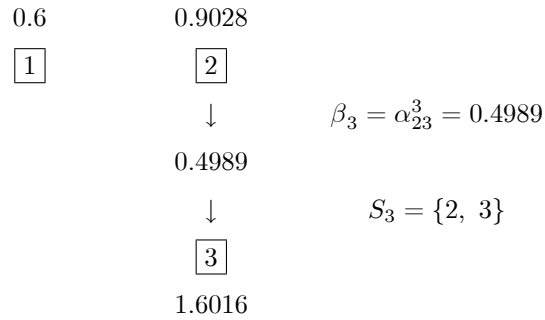
Tomamos  $\beta_1 = \alpha_{13}$  por ser el valor más pequeño distinto de cero de los  $\alpha_{ij}$ . El conjunto  $S_1$  está formado por los vértices unidos que contienen los vértices 1 y 3. Actualizamos los valores de todas las aristas de la siguiente forma  $\alpha_{ij}^2 = \alpha_{ij}^1 - \beta_1$ ,  $\forall i \in S_1$  y  $j \in S_1$ , con  $S_1 = \{1, 2, 3\}$ .

**l=2**



Tomamos  $\beta_2 = \alpha_{12}$  por ser el valor más pequeño distinto de cero de los  $\alpha_{ij}$ . El conjunto  $S_2$  esta formado por los vértices unidos que contienen los vértices 1 y 2. Actualizamos los valores de todas las aristas de la siguiente forma  $\alpha_{ij}^3 = \alpha_{ij}^2 - \beta_2$ ,  $\forall i \in S_2$  y  $j \in S_2$ , con  $S_2 = \{1, 2\}$ , las demás permanecen invariantes.

**l=3**



Tomamos  $\beta_3 = \alpha_{23}$  por ser el valor más pequeño distinto de cero de los  $\alpha_{ij}$ . El conjunto  $S_3$  esta formado por los vértices unidos que contienen los vértices 2 y 3. Actualizamos los valores de las aristas de la siguiente forma  $\alpha_{ij}^4 = \alpha_{ij}^3 - \beta_3$ ,  $\forall i \in S_3$  y  $j \in S_3$ , con  $S_3 = \{2, 3\}$ , las demás permanecen invariantes.

**1=4**

$$\begin{array}{cccc}
 0.6 & 0.4039 & & \\
 \boxed{1} & \boxed{2} & \beta_4 = 0.6 & S_4 = \{1\} \\
 & & \beta_5 = 0.4039 & S_5 = \{2\} \\
 & \boxed{3} & \beta_6 = 1.1027 & S_6 = \{3\} \\
 1.1027 & & & \tau = 6
 \end{array}$$

Puesto que todas las aristas tienen valor 0, termina el algoritmo eligiendo los  $\beta_l$ , por orden, igual al valor de los nodos.

Es decir, generamos variables Poisson  $Y_i$ , cuyos parámetros son los  $\beta_l$  que hemos calculado y tamaño igual al tamaño muestral ( $n = 25, 50, 100$ ).

$$Y = (\mathbf{Y}_1(0.4122), \mathbf{Y}_2(0.3337), \mathbf{Y}_3(0.4989), \mathbf{Y}_4(0.6), \mathbf{Y}_5(0.4039), \mathbf{Y}_6(1.1027))$$

Calculamos la matriz  $T$  con  $\tau = 6$  columnas donde ponemos unos en los lugares marcados por las diferentes  $S_\tau$  y 0 en el resto, quedando:

$$\mathbf{T} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

donde las variables Poisson correlacionadas las deseadas se obtienen haciendo

$$\mathbf{Z} = \mathbf{T} \times \mathbf{Y} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \mathbf{Y}_4 \\ \mathbf{Y}_5 \\ \mathbf{Y}_6 \end{pmatrix}$$

con lo que

$$\mathbf{Z} = (\mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_4, \mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_3 + \mathbf{Y}_5, \mathbf{Y}_1 + \mathbf{Y}_3 + \mathbf{Y}_6)$$

es una variable aleatoria Poisson tridimensional correlacionada.

La comparación de los diferentes procedimientos se lleva a cabo utilizando los mismos criterios de comparación que en la Sección 3.5.1 para los modelos con respuesta binaria.

La Tabla 3.16 muestra la salidas de la simulación en el caso en el que la estructura de la matriz de correlación de trabajo sea intercambiable con  $\alpha = 0.5$  en ambos grupos. Utilizando el criterio de la media vemos que los resultados que más se aproximan a los verdaderos valores de los parámetros los obtenemos con  $\lambda = -1/2$ . Los ECM menores los tenemos para el caso en que  $\lambda = 0$  y  $\lambda = 2/3$ . Estudiando el criterio de la probabilidad de cubrimiento observamos que los valores más próximos al 95 % los obtenemos cuando  $\lambda = 1$ .

La Tabla 3.17 muestra la salidas de la simulación en el caso en el que la estructura de la matriz de correlación de trabajo AR-1 con  $\alpha = 0.5$  en ambos grupos. Utilizando el criterio de la media vemos que los resultados que más se aproximan a los verdaderos valores de los parámetros los obtenemos con  $\lambda = -1/2$ . Los ECM menores los tenemos en el caso de que  $\lambda = 2/3$  y  $\lambda = 1$ . Estudiando el criterio de la probabilidad de cubrimiento observamos que los valores más próximos al 95 % los obtenemos cuando  $\lambda = 0, 1$  y  $2/3$ .

La Tabla 3.18 muestra la salidas de la simulación en el caso en el que la estructura de la matriz de correlación de trabajo desestructurada con coeficientes de correlación 0.5, 0.1 y 0.3 en ambos grupos. Utilizando el criterio de la media vemos que los resultados que más se aproximan a los verdaderos valores de los parámetros los obtenemos con  $\lambda = -1/2$ . Los ECM se observan para el caso en el que  $\lambda = -1/2, \lambda = 0$  y  $\lambda = 2/3$ . Estudiando el criterio de la probabilidad de cubrimiento observamos que los valores más próximos al 95 % los obtenemos cuando  $\lambda = 1, 0$  y  $2/3$ .

Por lo que podríamos concluir, para casi todas las estructuras de la matriz de correlación de trabajo y para todos los tamaños muestrales la mejor alternativa se obtiene para el procedimiento de estimación correspondiente a  $\lambda = 2/3$ .

Tabla 3.16. Resultados simulación para respuesta Poisson con matriz de correlación de trabajo intercambiable  $\alpha=0.5$  en ambos grupos

$n$	$\text{Parámetros}$	$\text{Estadístico}$	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
100	$\beta_0$	<i>Media</i>	0.0978	0.0968	0.0963	0.0961	0.0953
		<i>ECM</i>	0.1139	0.1131	0.1129	0.1130	0.1133
		<i>Pr ob Cub</i>	0.9410	0.9440	0.9450	0.9455	0.9420
	$\beta_1$	<i>Media</i>	0.1997	0.2007	0.2012	0.2014	0.2021
		<i>ECM</i>	0.1514	0.1502	0.1500	0.1500	0.1504
		<i>Pr ob Cub</i>	0.9330	0.9385	0.9380	0.9385	0.9380
	$\beta_2$	<i>Media</i>	0.2009	0.2012	0.2014	0.2015	0.2018
		<i>ECM</i>	0.0410	0.0405	0.0405	0.0405	0.0406
		<i>Pr ob Cub</i>	0.9385	0.9425	0.9410	0.9410	0.9410
	$\beta_3$	<i>Media</i>	0.0997	0.0993	0.0992	0.0991	0.0988
		<i>ECM</i>	0.0533	0.0527	0.0526	0.0526	0.0528
		<i>Pr ob Cub</i>	0.9370	0.9410	0.9415	0.9420	0.9405
50	$\beta_0$	<i>Media</i>	0.0956	0.0942	0.0936	0.0933	0.0921
		<i>ECM</i>	0.1603	0.1595	0.1595	0.1597	0.1600
		<i>Pr ob Cub</i>	0.9425	0.9440	0.9440	0.9445	0.9445
	$\beta_1$	<i>Media</i>	0.2005	0.2022	0.2028	0.2029	0.2039
		<i>ECM</i>	0.2078	0.2069	0.2069	0.2072	0.2074
		<i>Pr ob Cub</i>	0.9485	0.9530	0.9520	0.9505	0.9510
	$\beta_2$	<i>Media</i>	0.2006	0.2012	0.2014	0.2014	0.2018
		<i>ECM</i>	0.0577	0.0572	0.0572	0.0573	0.0576
		<i>Pr ob Cub</i>	0.9340	0.9365	0.9375	0.9365	0.9345
	$\beta_3$	<i>Media</i>	0.1003	0.0997	0.0994	0.0993	0.0989
		<i>ECM</i>	0.0740	0.0736	0.0736	0.0735	0.0738
		<i>Pr ob Cub</i>	0.9395	0.9435	0.9450	0.9440	0.9420
25	$\beta_0$	<i>Media</i>	0.0959	0.0934	0.0923	0.0919	0.0896
		<i>ECM</i>	0.2261	0.2259	0.2261	0.2263	0.2271
		<i>Pr ob Cub</i>	0.9295	0.9295	0.9295	0.9285	0.9260
	$\beta_1$	<i>Media</i>	0.1871	0.1898	0.1910	0.1915	0.1931
		<i>ECM</i>	0.3046	0.3040	0.3042	0.3043	0.3056
		<i>Pr ob Cub</i>	0.9290	0.9285	0.9290	0.9290	0.9250
	$\beta_2$	<i>Media</i>	0.2004	0.2014	0.2018	0.2020	0.2021
		<i>ECM</i>	0.0813	0.0810	0.0810	0.0810	0.0820
		<i>Pr ob Cub</i>	0.9230	0.9270	0.9270	0.9270	0.9235
	$\beta_3$	<i>Media</i>	0.1021	0.1010	0.1006	0.1004	0.0996
		<i>ECM</i>	0.1053	0.1050	0.1050	0.1051	0.1056
		<i>Pr ob Cub</i>	0.9335	0.9355	0.9350	0.9355	0.9315

Tabla 3.17. Resultados simulación para respuesta Poisson con matriz  
de correlación de trabajo AR-1 con  $\alpha=0.5$  en ambos grupos

$n$	$\text{Parámetros}$	$\text{Estadístico}$	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
100	$\beta_0$	<i>Media</i>	0.0944	0.0935	0.0928	0.0930	0.0922
		<i>ECM</i>	0.1247	0.1236	0.1234	0.1234	0.1239
		<i>Pr ob Cub</i>	0.9380	0.9410	0.9385	0.9410	0.9405
	$\beta_1$	<i>Media</i>	0.2037	0.2044	0.2043	0.2049	0.2057
		<i>ECM</i>	0.1629	0.1618	0.1617	0.1617	0.1620
		<i>Pr ob Cub</i>	0.9410	0.9455	0.9450	0.9475	0.9445
	$\beta_2$	<i>Media</i>	0.2016	0.2019	0.2016	0.2021	0.2024
		<i>ECM</i>	0.0490	0.0484	0.0492	0.0483	0.0486
		<i>Pr ob Cub</i>	0.9345	0.9420	0.9420	0.9435	0.9390
	$\beta_3$	<i>Media</i>	0.0987	0.0985	0.0980	0.0983	0.0979
		<i>ECM</i>	0.0629	0.0621	0.0621	0.0620	0.0623
		<i>Pr ob Cub</i>	0.9400	0.9470	0.9440	0.9465	0.9465
50	$\beta_0$	<i>Media</i>	0.0895	0.0886	0.0882	0.0880	0.0868
		<i>ECM</i>	0.1772	0.1768	0.1769	0.1770	0.1781
		<i>Pr ob Cub</i>	0.9395	0.9435	0.9445	0.9435	0.9400
	$\beta_1$	<i>Media</i>	0.2028	0.2039	0.2044	0.2046	0.2054
		<i>ECM</i>	0.2290	0.2283	0.2283	0.2284	0.2290
		<i>Pr ob Cub</i>	0.9505	0.9505	0.9505	0.9505	0.9495
	$\beta_2$	<i>Media</i>	0.2024	0.2027	0.2029	0.2029	0.2033
		<i>ECM</i>	0.0700	0.0697	0.0696	0.0696	0.0702
		<i>Pr ob Cub</i>	0.9320	0.9370	0.9365	0.9365	0.9325
	$\beta_3$	<i>Media</i>	0.0997	0.0992	0.0991	0.0990	0.0986
		<i>ECM</i>	0.0905	0.0899	0.0898	0.0898	0.0902
		<i>Pr ob Cub</i>	0.9350	0.9405	0.9415	0.9410	0.9380
25	$\beta_0$	<i>Media</i>	0.0906	0.0889	0.0881	0.0877	0.0844
		<i>ECM</i>	0.2462	0.2459	0.2460	0.2461	0.2457
		<i>Pr ob Cub</i>	0.9340	0.9355	0.9360	0.9360	0.9295
	$\beta_1$	<i>Media</i>	0.1919	0.1938	0.1946	0.1950	0.1964
		<i>ECM</i>	0.3360	0.3356	0.3357	0.3357	0.3351
		<i>Pr ob Cub</i>	0.9320	0.9340	0.9335	0.9335	0.9270
	$\beta_2$	<i>Media</i>	0.2017	0.2024	0.2027	0.2028	0.2023
		<i>ECM</i>	0.0989	0.0983	0.0982	0.0982	0.0994
		<i>Pr ob Cub</i>	0.9305	0.9330	0.9345	0.9345	0.9240
	$\beta_3$	<i>Media</i>	0.1019	0.1011	0.1008	0.1007	0.0997
		<i>ECM</i>	0.1306	0.1299	0.1298	0.1298	0.1299
		<i>Pr ob Cub</i>	0.9300	0.9315	0.9310	0.9310	0.9240

Tabla 3.18. Resultados de simulación para respuesta Poisson con matriz de correlación de trabajo desestructurada con coeficientes de correlación 0.5, 0.1 y 0.3 en ambos grupos

$n$	$\text{Parámetros}$	$\text{Estadístico}$	$\lambda = -1/2$	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$
100	$\beta_0$	<i>Media</i>	0.0978	0.0968	0.0964	0.0962	0.0952
		<i>ECM</i>	0.1316	0.1313	0.1314	0.1314	0.1314
		<i>Pr ob Cub</i>	0.9440	0.9455	0.9460	0.9470	0.9465
	$\beta_1$	<i>Media</i>	0.1999	0.2008	0.2012	0.2014	0.2024
		<i>ECM</i>	0.1754	0.1749	0.1751	0.1751	0.1752
		<i>Pr ob Cub</i>	0.9365	0.9420	0.9430	0.9415	0.9370
	$\beta_2$	<i>Media</i>	0.1997	0.2000	0.2002	0.2003	0.2007
		<i>ECM</i>	0.0531	0.0529	0.0529	0.0529	0.0529
		<i>Pr ob Cub</i>	0.9425	0.9475	0.9475	0.9480	0.9475
	$\beta_3$	<i>Media</i>	0.1009	0.1006	0.1005	0.1004	0.1000
		<i>ECM</i>	0.0698	0.0696	0.0696	0.0696	0.0697
		<i>Pr ob Cub</i>	0.9430	0.9485	0.9480	0.9475	0.9460
50	$\beta_0$	<i>Media</i>	0.0969	0.0953	0.0957	0.0951	0.0941
		<i>ECM</i>	0.1843	0.1848	0.1846	0.1849	0.1853
		<i>ProbCub</i>	0.9430	0.9445	0.9440	0.9445	0.9430
	$\beta_1$	<i>Media</i>	0.1981	0.2003	0.1996	0.2005	0.2013
		<i>ECM</i>	0.2376	0.2382	0.2377	0.2382	0.2385
		<i>Pr ob Cub</i>	0.9535	0.9555	0.9560	0.9555	0.9540
	$\beta_2$	<i>Media</i>	0.1989	0.1996	0.1994	0.1996	0.2000
		<i>ECM</i>	0.0738	0.0738	0.0737	0.0738	0.0740
		<i>Pr ob Cub</i>	0.9440	0.9475	0.9485	0.9480	0.9470
	$\beta_3$	<i>Media</i>	0.1017	0.1008	0.1011	0.1007	0.1004
		<i>ECM</i>	0.0961	0.0961	0.0960	0.0961	0.0963
		<i>Pr ob Cub</i>	0.9460	0.9500	0.9495	0.9500	0.9490
25	$\beta_0$	<i>Media</i>	0.0908	0.0884	0.0875	0.0871	0.0844
		<i>ECM</i>	0.2604	0.2613	0.2617	0.2618	0.2643
		<i>Pr ob Cub</i>	0.9405	0.9405	0.9405	0.9400	0.9385
	$\beta_1$	<i>Media</i>	0.2011	0.2039	0.2044	0.2051	0.2073
		<i>ECM</i>	0.3486	0.3504	0.3509	0.3512	0.3526
		<i>Pr ob Cub</i>	0.9415	0.9415	0.9415	0.9420	0.9410
	$\beta_2$	<i>Media</i>	0.2007	0.2018	0.2021	0.2022	0.2030
		<i>ECM</i>	0.1051	0.1052	0.1052	0.1052	0.1061
		<i>Pr ob Cub</i>	0.9330	0.9330	0.9335	0.9335	0.9295
	$\beta_3$	<i>Media</i>	0.0986	0.0975	0.0974	0.0971	0.0960
		<i>ECM</i>	0.1387	0.1391	0.1391	0.1392	0.1395
		<i>Pr ob Cub</i>	0.9350	0.9340	0.9340	0.9340	0.9320





## Capítulo 4

# Medidas de Diagnóstico

### 4.1. Introducción

El objetivo de las medidas de diagnóstico es identificar subconjuntos de datos que parezcan tener una influencia desproporcionada en la estimación del modelo y establecer que partes del modelo estimado se ven más afectadas por estos subconjuntos. Se debe estudiar el que un pequeño subconjunto de datos pueda tener una influencia desproporcionada en la estimación de los parámetros o en las predicciones ya que, si este es el caso, es bastante probable que los modelos estimados estén basados principalmente en estos subconjuntos en lugar de en la mayoría de los datos.

Tan et al. [107] proponen ciertas medidas de diagnóstico para chequear la conveniencia de los modelos de regresión marginal pero sólo en el análisis de datos binarios correlacionados. Venezuela et al. [110] presentaron una extensión de estas medidas válidas para el análisis de regresión de medidas repetidas con las EEG en general. En particular, consideran la matriz “hat”, la distancia de Cook, residuos estandarizados y gráficos de probabilidad empírica (ver, Atkinson [10]).

Chang [25], entre otros, presenta un test no paramétrico para chequear posibles patrones de no-aleatoriedad. Pan [78] propuso una versión modificada (QIC) del criterio de información de Akaike que funciona bien para la selección de variables y de la matriz de correlación de trabajo. Preisser y Qaqish [91] propusieron diagnósticos para las EEG

que consisten en comparar diferentes características entre el modelo considerando todas las observaciones y el modelo presentado eliminando bien una observación bien todas las observaciones pertenecientes a un individuo.

Algunos de los diagnósticos para las ecuaciones de estimación generalizadas son generalizaciones de éstos para el modelo lineal generalizado (Pregibon [90], Williams [114], McCullagh y Nelder [68]), y éstos se reducen a conocidas medidas de influencia en la regresión lineal que se puede encontrar en Cook y Weisberg [27], Belsley, Kuh y Welsch [18], y un trabajo interesante con comentarios de Chatterjee y Hadi [26]. Estos autores consideran que las medidas de influencia pueden ser clasificadas dentro de cinco grupos:

1. Medidas basadas en residuos.
2. Medidas basadas en la matriz de predicción.
3. Medidas basadas en el volumen de las elipsoides de confianza.
4. Medidas basadas en funciones de influencia.
5. Medidas basadas en influencia parcial.

En relación al primer grupo de medidas Oh et al. [76] proponen gráficos de residuos basados en gráficos cuantil-cuantil de una distribución  $\chi^2$  para analizar datos con medidas repetidas discretas. Estos autores consideran no sólo los residuos de Pearson si no también los de Anscombe y desviación. En la Sección 4.2, extendemos el trabajo de Oh et al. [76] considerando una familia de gráficos de residuos basada en la familia de residuos  $c_{ij}^\phi$  definida en (3.5). Además, se lleva a cabo un estudio de simulación en la Sección 4.3 para comparar los diferentes gráficos propuestos.

Por otra parte, las medidas basadas en el volumen de elipsoides de confianza no han sido generalizadas. Estas nuevas medidas de diagnóstico se definen en la Sección 4.4. Ilustramos el uso de las nuevas medidas de diagnóstico en la Sección 4.5 con unos ejemplos.

## 4.2. Gráficos cuantil-cuantil basados en residuos $\phi$ -divergencia

Oh et al. [76] propusieron unas gráficas de residuos para investigar la bondad de ajuste para datos con medidas repetidas utilizando las EEG. Estos gráficos se basan en gráficos cuantil-cuantil de una distribución  $\chi^2$  y utilizan no sólo los residuos de Pearson sino también los de Anscombe y desviación. En esta Sección, extendemos estos gráficos considerando la familia de residuos  $c_{ij}^\phi$  definida en (3.5) y (3.6). De esta forma, al igual que en el capítulo anterior, se había conseguido una mejora en la estimación de los parámetros utilizando unos residuos distintos a los de Pearson en el procedimiento de las EEG, se espera con la extensión de estos gráficos unas medidas de diagnóstico más eficientes.

En modelos de regresión univariante, si un modelo se ajusta bien a los datos, se espera que los residuos se aproximen a una distribución normal. Si suponemos que las estructuras de la media y de la covarianza son correctas, entonces se espera también que el vector de residuos tenga distribución normal con vector de medias cero. Utilizando esta idea, Park y Lee [86] propusieron un gráfico simple de residuos para chequear la bondad del ajuste para datos con medidas repetidas distribuidas como una normal. Oh et al. [76] extendieron este gráfico de residuos para estudiar la bondad del ajuste de las EEG, basándose en residuos de Pearson, Anscombe y desviación. Nosotros extendemos las gráficas de residuos de Oh et al. [76] utilizando los residuos basados en una familia de divergencias que tienen como casos particulares los residuos de Pearson y de desviación.

Sea  $\mathbf{c}_i^\phi$  el vector de residuos  $t_i \times 1$  del individuo  $i$ -ésimo con coordenada  $j$ -ésima,  $c_{ij}^\phi$ . Sea

$$\mathbf{M}_i = \mathbf{I} - \mathbf{H}_i$$

donde

$$\mathbf{H}_i = \mathbf{W}_i^{\frac{1}{2}} \mathbf{X}_i \mathbf{J}_1 \mathbf{X}_i^T \mathbf{W}_i^{\frac{1}{2}} \quad (4.1)$$

con

$$\mathbf{W}_i = \mathbf{\Lambda}_i \mathbf{V}_i^{-1} \mathbf{\Lambda}_i, \quad \mathbf{\Lambda}_i = \text{diag} \left( \frac{d\boldsymbol{\mu}_i}{d\boldsymbol{\eta}_i^T} \right)$$

y

$$\mathbf{J}_1 = \left[ \sum_{i=1}^n \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^T \hat{\mathbf{V}}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right]^{-1} \quad (4.2)$$

la matriz de covarianzas asintótica de  $\mathbf{c}_i^\phi$ . Cuando conocemos  $\mathbf{M}_i$  y el modelo para la media es correcto,  $q_i^\phi = \left(\mathbf{c}_i^\phi\right)^T \mathbf{M}_i^{-1} \mathbf{c}_i^\phi$ ,  $i = 1, \dots, n$ , son distribuidos aproximadamente como una distribución  $\chi^2$  con  $t_i$  grados de libertad. Es decir,

$$q_i^\phi = \left(\mathbf{c}_i^\phi\right)^T \mathbf{M}_i^{-1} \mathbf{c}_i^\phi \sim \chi^2(t_i).$$

Cuando el número de respuestas de un mismo individuo es igual a  $t$  para todo  $i$ , podemos construir fácilmente un gráfico cuantil-cuantil con los  $q_i^\phi$  observados utilizando la distribución  $\chi^2$ . Sean  $q_{(1)}^\phi \leq \dots \leq q_{(n)}^\phi$  valores ordenados de los  $q_i^\phi$ . Entonces,  $q_i^\phi$ , es de hecho el percentil empírico  $100 \times \frac{i}{n}$ , y de la distribución  $\chi^2$  con  $t$  grados de libertad, podemos obtener los correspondientes cuantiles  $\psi_{(1)} \leq \dots \leq \psi_{(n)}$ . Entonces, el gráfico cuantil-cuantil es el gráfico de  $(q_{(i)}, \psi_{(i)})$  desde el que podemos estudiar el ajuste del modelo y la identificación de “outlying”.

Cuando el número de respuestas  $t_i$  difiere de individuo a individuo tendremos observaciones incompletas, los grados de libertad difieren de individuo a individuo, y no es posible representar un gráfico cuantil-cuantil basado en la distribución  $\chi^2$ . En este caso, proponemos el uso de gráfico cuantil-cuantil al que le hacemos una simple transformación propuesta por Hawkins y Wixley [49] para conseguir que se aproximen a la normal, de forma que

$$q_{iN}^\phi = \frac{\left(q_i^\phi\right)^{1/4} - (t_i - 0.5)^{1/4}}{(8\sqrt{t_i})^{-1/2}}, \quad i = 1, \dots, n,$$

se distribuyen aproximadamente como una distribución normal estándar  $N(0, 1)$ . Aquí, la media y la varianza son aproximaciones calculadas con el desarrollo en serie de Taylor de la expresión del logaritmo del primer y segundo momento. Park y Lee [86] demostraron que esta transformación funciona bien para medidas repetidas normalmente distribuidas.

### 4.3. Estudio de Simulación

Para estudiar el comportamiento de las gráficas cuantil-cuantil de residuos en variables discretas, vamos a llevar a cabo un estudio de simulación primero para variables con respuesta binaria y a continuación para variables con respuesta Poisson.

Consideramos medidas repetidas tomadas en 4 tiempos diferentes ( $t=4$ ) obtenidas de dos grupos donde las variables respuesta son variables binarias correlacionadas o variables Poisson correlacionadas. A partir de un modelo marginal generamos un total de 100 muestras, para cada uno de los cuatro modelos considerados.

Se consideran las gráficas cuantil-cuantil basadas en  $q_i^{\phi(\lambda)}$  donde  $\phi(\lambda)$  viene definido en (2.28) y se consideran los valores de  $\lambda$ ,  $\lambda = -1/2, 0, 2/3, 1$ . El eje  $x$  es el de los cuantiles esperados procedentes de la distribución  $\chi^2$  con cuatro grados de libertad y el eje  $y$  es el de los cuantiles observados de  $q_{(1)}^{\phi(\lambda)}, \dots, q_{(50)}^{\phi(\lambda)}$ . Para cada valor de  $x$ , hay dos valores para la  $y$ , que son el máximo y el mínimo de los 100 cuantiles simulados.

También se presentan los gráficos cuantil-cuantil considerando en el eje  $y$  las medias de los 100 cuantiles simulados.

#### 4.3.1. Para variables binarias

Consideramos medidas repetidas obtenidas de dos grupos de 25 individuos cada una donde las variables respuesta son variables binarias correlacionadas. Se supone que el modelo marginal viene dado por:

$$\text{logit}(\mu_{ij}) = \log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \beta_0 + \beta_1 x_i + \beta_2 j + \beta_3 x_i j \quad (4.3)$$

donde

$$x_i = \begin{cases} 0 & \text{grupo 1} \\ 1 & \text{grupo 2} \end{cases} \quad y \quad j = 1, 2, 3.$$

Consideramos los modelos:

$$\begin{aligned} \text{Modelo 1: } & \text{logit}(\mu_{ij}) = \beta_0 \\ \text{Modelo 2: } & \text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_i \\ \text{Modelo 3: } & \text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j \\ \text{Modelo 4: } & \text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j + \beta_3 x_i j \end{aligned} \quad (4.4)$$

Los verdaderos valores de los parámetros elegidos son

$$(\beta_0, \beta_1, \beta_2, \beta_3) = (-1, 0.2, 0.2, 1).$$

Suponemos que no hay sobredispersión por lo que valor del parámetro  $\gamma = 1$ . La estructura de la matriz de correlación de trabajo que se considera es la intercambiable con  $\rho = 0.5$ .

Generamos un total de 100 muestras con variables binarias correlacionadas, de acuerdo al algoritmo presentado en la Sección 3.5.2, para cada uno de los cuatro modelos considerados.

Las Figuras 4.1-4.4 muestran los gráficos cuantil-cuantil para la media de 100 cuantiles simulados para los 4 modelos definidos en (4.4), respectivamente. En las Figuras 4.5-4.8 se presenta los rangos de los  $q_i^{\phi(\lambda)}$  para  $\lambda = -1/2, 0, 2/3$  y 1 basados en gráficos cuantil-cuantil. En el eje  $x$  están los cuantiles esperados de una distribución  $\chi^2$  con 4 grados de libertad y en el eje  $y$  los cuantiles observados de  $q_{(1)}^{\phi(\lambda)}, \dots, q_{(50)}^{\phi(\lambda)}$  obtenidos de cada modelo. Para cada valor  $x$  hay dos valores  $y$ : el máximo y el mínimo de los 100 cuantiles simulados.

Teniendo en cuenta que el mejor procedimiento de diagnóstico es aquel en que la nube de puntos en el gráfico se aproxima lo máximo posible a la recta  $y = x$  cuando el modelo ajustado es bueno y se aleja lo máximo posible cuando el modelo ajustado se aleja del modelo verdadero. En la Figura 4.4 que es la correspondiente al verdadero modelo, el Modelo 4, el mejor procedimiento es el correspondiente al gráfico cuantil-cuantil basado en  $q_i^{\phi(2/3)}$  seguido de los procedimientos basados en  $q_i^{\phi(0)}$  (residuo de desviación) y  $q_i^{\phi(1)}$  (residuos de Pearson) que tienen un comportamiento análogo. Sin embargo, el gráfico cuantil-cuantil basado en  $q_i^{\phi(1)}$  queda descartado si observamos la Figura 4.8 ya que su rango de variación es demasiado amplio. Por último, el gráfico cuantil-cuantil correspondiente a  $q_i^{\phi(-1/2)}$  es el que peor comportamiento tiene ya que aunque no detecta ningún “outlying” toda la nube de puntos está separada de la recta  $y = x$ . Este procedimiento aunque se muestra en el resto de figuras queda descartado.

En cuanto a los Modelos 2 y 3, las Figuras 4.2 y 4.3 muestran como mejores procedimientos las gráficas correspondientes a  $q_i^{\phi(0)}$  y  $q_i^{\phi(2/3)}$ . Conclusión que viene reforzada por las Figuras 4.6 y 4.7. Por tanto, se recomienda el uso del gráfico cuantil-cuantil basado en  $q_i^{\phi(2/3)}$  como medida de diagnóstico para detectar observaciones “outlying” y verificar la bondad del ajuste del modelo.

Figura 4.1. Gráfico cuantil-cuantil para la media de 100 cuantiles simulados para el Modelo 1

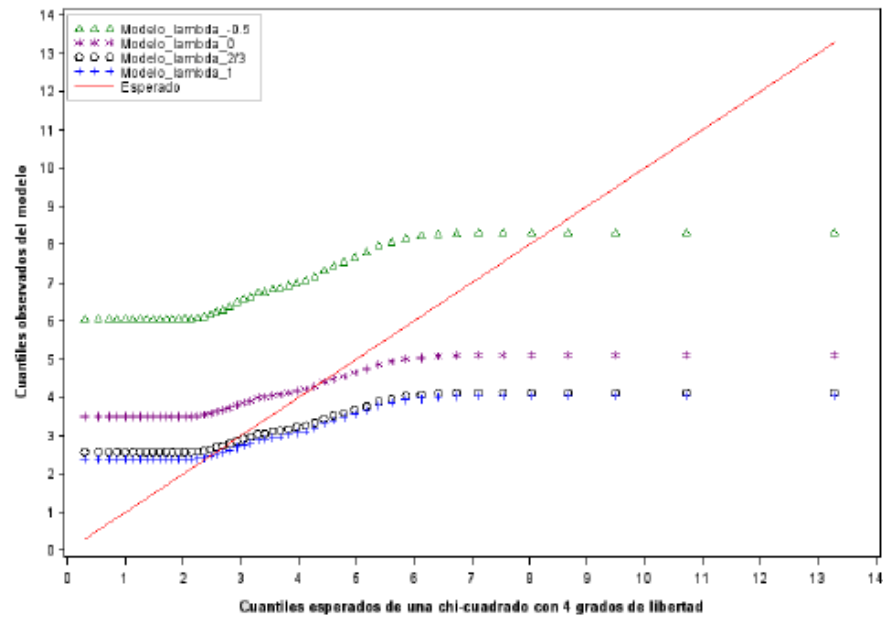


Figura 4.2. Gráfico cuantil-cuantil para la media de 100 cuantiles simulados para el Modelo 2

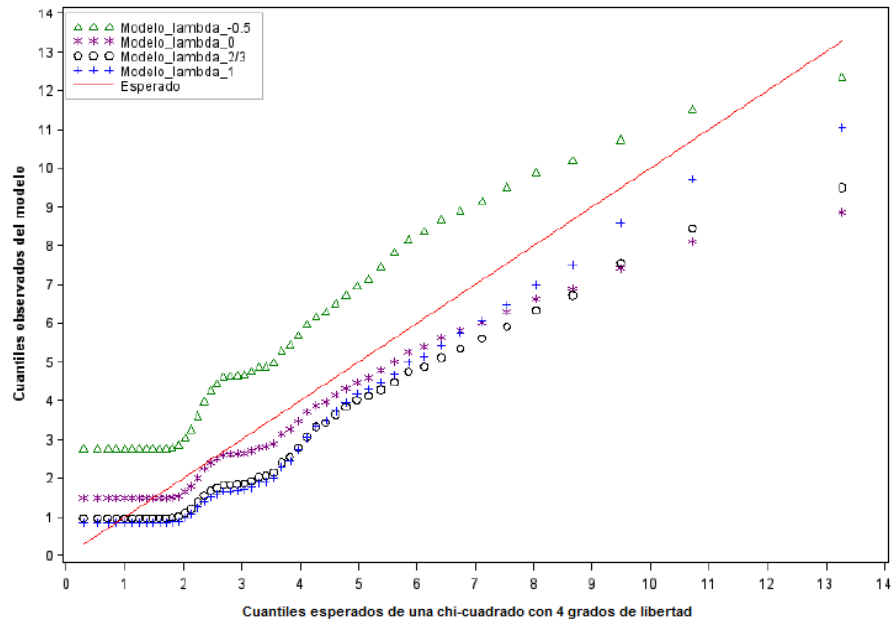




Figura 4.3. Gráfico cuantil-cuantil para la media de 100 cuantiles simulados para el Modelo 3

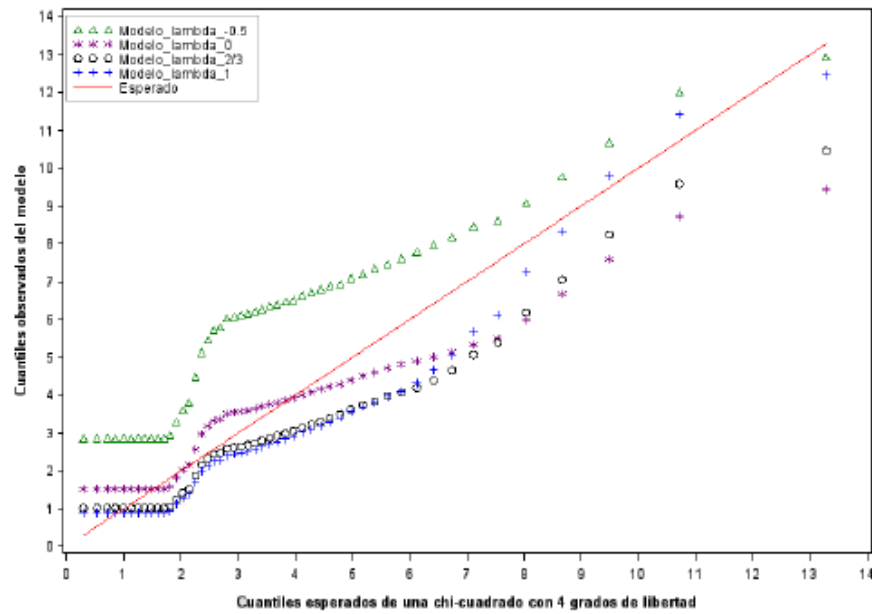


Figura 4.4. Gráfico cuantil-cuantil para la media de 100 cuantiles simulados para el Modelo 4

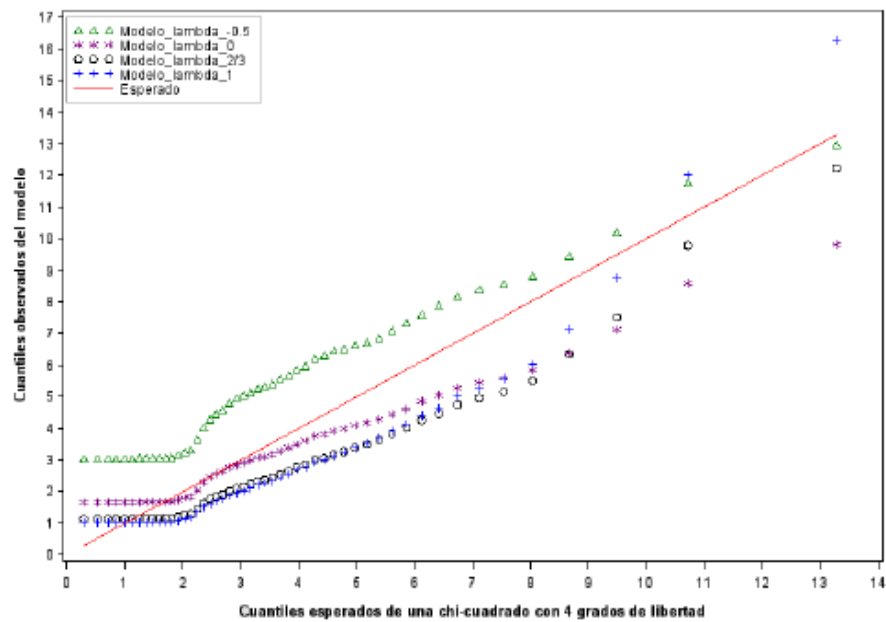


Figura 4.5. Gráfico de la banda min-max para 100 cuantiles simulados para el Modelo 1

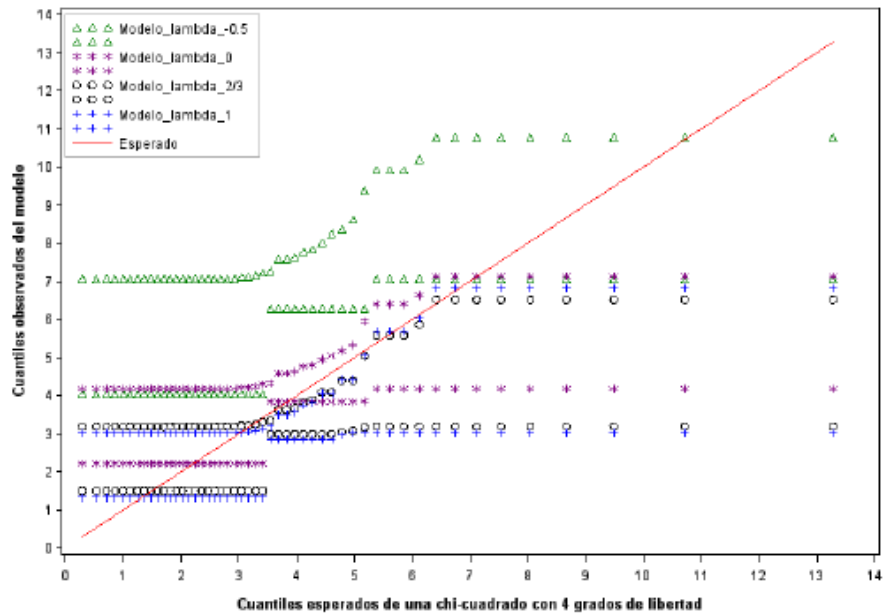


Figura 4.6. Gráfico de la banda min-max para 100 cuantiles simulados para el Modelo 2

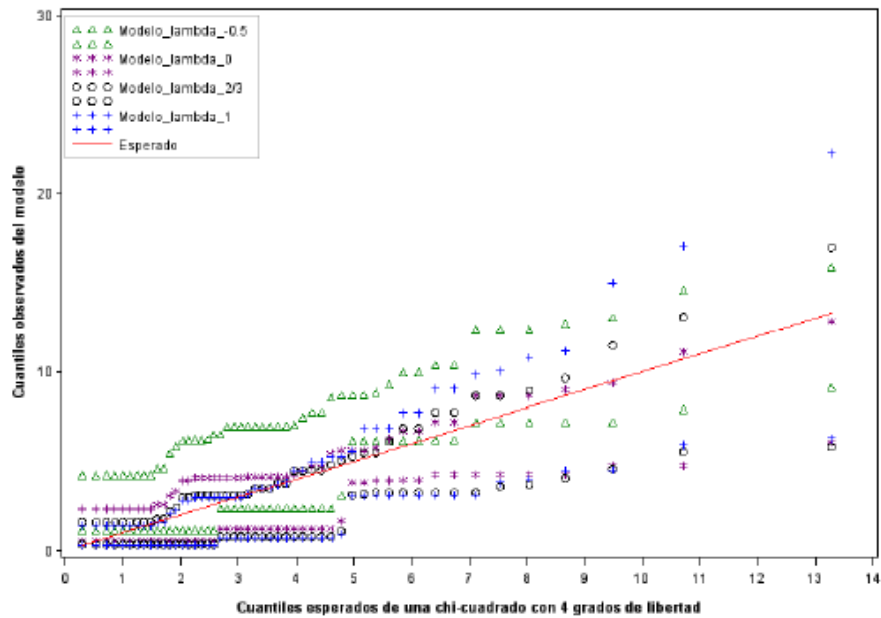


Figura 4.7. Gráfico de la banda min-max para 100 cuantiles simulados para el Modelo 3

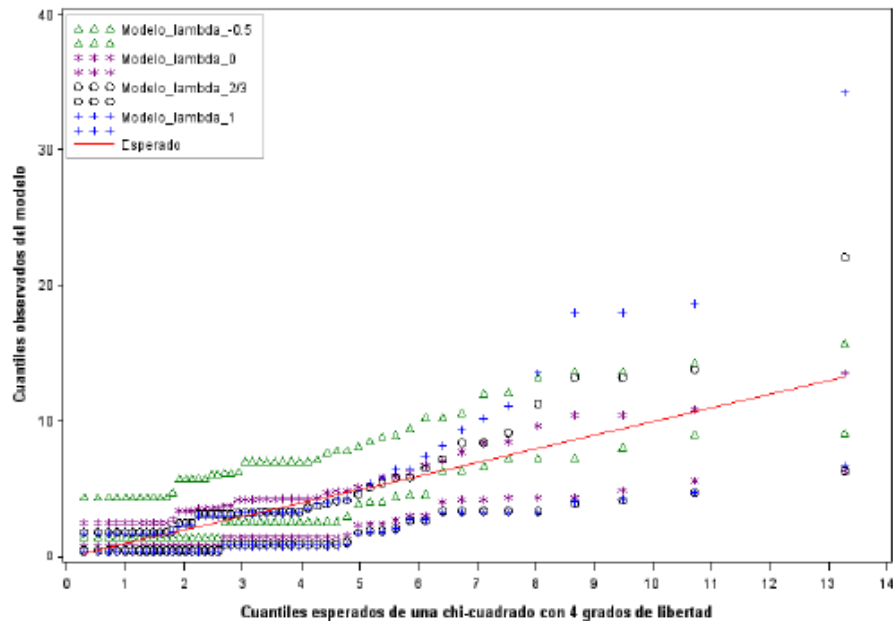
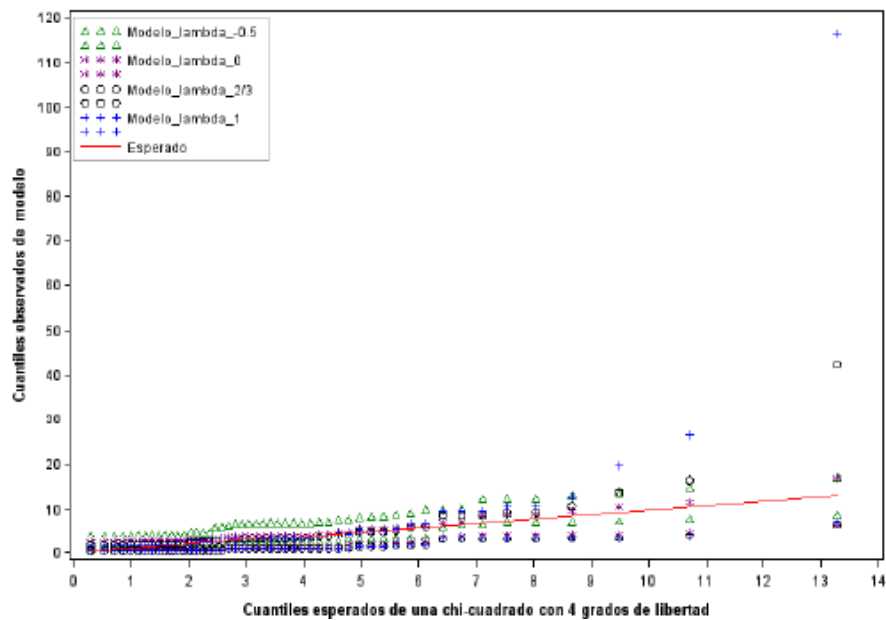


Figura 4.8. Gráfico de la banda min-max para 100 cuantiles simulados para el Modelo 4



### 4.3.2. Para variables Poisson

Consideramos medidas repetidas obtenidas de dos grupos de 25 individuos cada una donde las variables respuesta son variables Poisson correlacionadas. Se supone que el verdadero modelo marginal viene dado por:

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j + \beta_3 x_i j \quad (4.5)$$

donde

$$x_i = \begin{cases} 0 & \text{grupo 1} \\ 1 & \text{grupo 2} \end{cases} \quad y \quad j = 1, 2, 3.$$

Consideramos los modelos:

$$\begin{aligned} \text{Modelo 1: } & \log(\mu_{ij}) = \beta_0 \\ \text{Modelo 2: } & \log(\mu_{ij}) = \beta_0 + \beta_1 x_i \\ \text{Modelo 3: } & \log(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j \\ \text{Modelo 4: } & \log(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j + \beta_3 x_i j \end{aligned} \quad (4.6)$$

Los verdaderos valores de los parámetros elegidos son

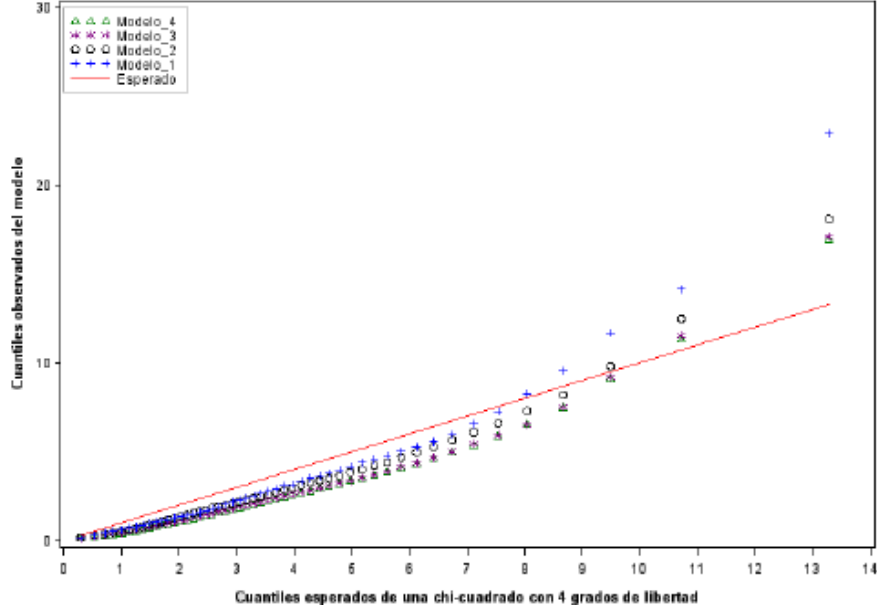
$$(\beta_0, \beta_1, \beta_2, \beta_3) = (0.1, 0.2, 0.2, 0.1).$$

Suponemos que no hay sobredispersión por lo que valor del parámetro  $\gamma = 1$ . Consideramos la estructura de la matriz de correlación de trabajo intercambiable con  $\rho = 0.5$ .

Generamos un total de 100 muestras con variables Poisson correlacionadas, de acuerdo al algoritmo presentado en la Sección 3.6.2 bajo el Modelo 4 que coincide con el modelo dado en (4.5).

En la Figura 4.9 se presenta el gráfico cuantil-cuantil basado en los residuos de Pearson para la media de 100 cuantiles simulados para cada uno de los modelos dados en (4.6). Como es de esperar el mejor modelo que se ajusta es el Modelo 4 que es el verdadero.

Figura 4.9. Gráfico cuantil-cuantil basado en el residuo de Pearson para los Modelos dados en (4.6)



Las Figuras 4.10-4.13 muestran los gráficos cuantil-cuantil para la media de 100 cuantiles simulados para los 4 modelos definidos en (4.6), respectivamente. En las Figuras 4.14-4.17 se presenta los rangos de los  $q_i^{\phi(\lambda)}$  para  $\lambda = -1/2, 0, 2/3$  y 1 basados en gráficos cuantil-cuantil. En el eje  $x$  están los cuantiles esperados de una distribución  $\chi^2$  con 4 grados de libertad y en el eje  $y$  los cuantiles observados de  $q_{(1)}^{\phi(\lambda)}, \dots, q_{(50)}^{\phi(\lambda)}$  obtenidos de cada modelo. Para cada valor  $x$  hay dos valores  $y$ : el máximo y el mínimo de los 100 cuantiles simulados.

De estas figuras se deduce lo siguiente: Para el verdadero modelo, el Modelo 4, el mejor procedimiento es el correspondiente al gráfico cuantil-cuantil basado en  $q_i^{\phi(0)}$  (residuo de desviación) seguido de cerca por el basado en  $q_i^{\phi(2/3)}$  y en tercer lugar el basado en  $q_i^{\phi(1)}$  (residuo de Pearson). Ya que se puede ver en la Figura 4.13 que son en ese orden los más próximos a la recta  $y = x$ . En la Figura 4.17 son los que tienen un rango más ajustado en torno a la recta  $y = x$ . Para el resto de modelos, Modelo 1, 2 y 3, los mejores procedimientos son en orden decreciente de mejor a peor los gráficos cuantil-

cuantil correspondientes a  $q_i^{\phi_{(1)}}$ ,  $q_i^{\phi_{(2/3)}}$  y  $q_i^{\phi_{(0)}}$ . Ya que como se puede ver en los gráficos 4.10-4.12 son los más alejados de la recta  $y = x$ . Esta conclusión está de acuerdo con la que se obtiene a partir de las Figuras 4.14-4.16. Por tanto, el gráfico cuantil-cuantil basado en  $q_i^{\phi_{(2/3)}}$  es el mejor compromiso entre mostrar un buen ajuste para el verdadero modelo y la detección de modelos no verdaderos. En cuanto el gráfico cuantil-cuantil basado en  $q_i^{\phi_{(-1/2)}}$  claramente es bastante malo. Basta con observar las Figuras 4.13 y 4.17 que se corresponden con el modelo verdadero.

Figura 4.10. Gráfico cuantil-cuantil para la media de 100 cuantiles simulados para el Modelo 1

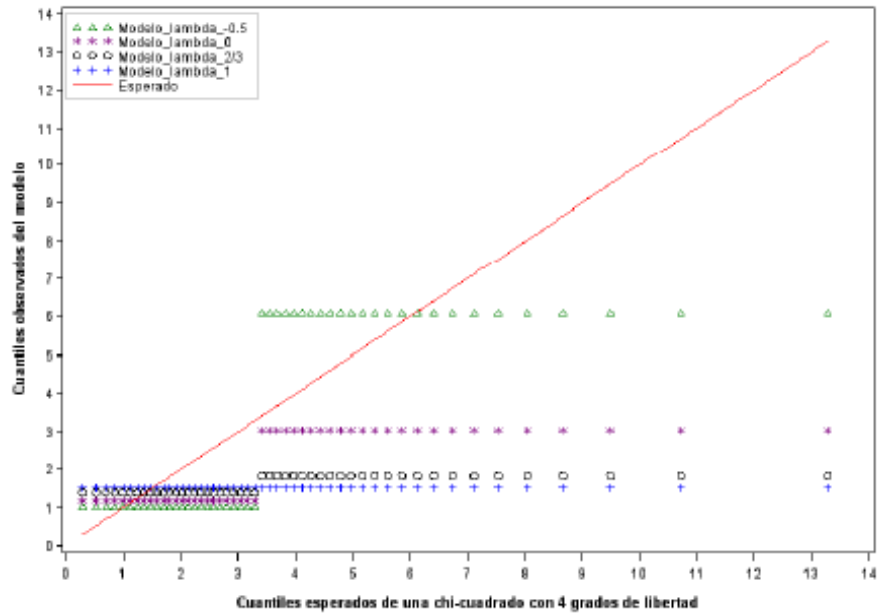


Figura 4.11. Gráfico cuantil-cuantil para la media de 100 cuantiles simulados para el Modelo 2

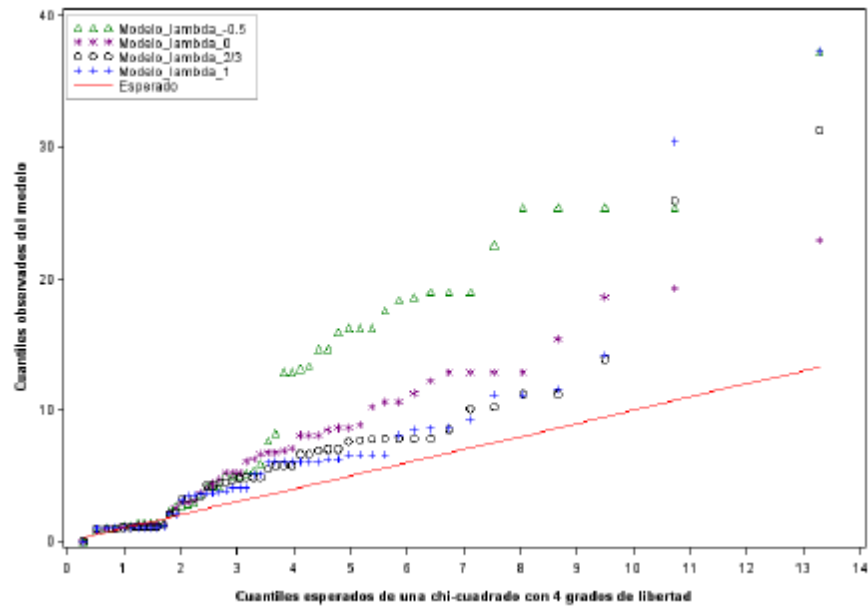


Figura 4.12. Gráfico cuantil-cuantil para la media de 100 cuantiles simulados para el Modelo 3

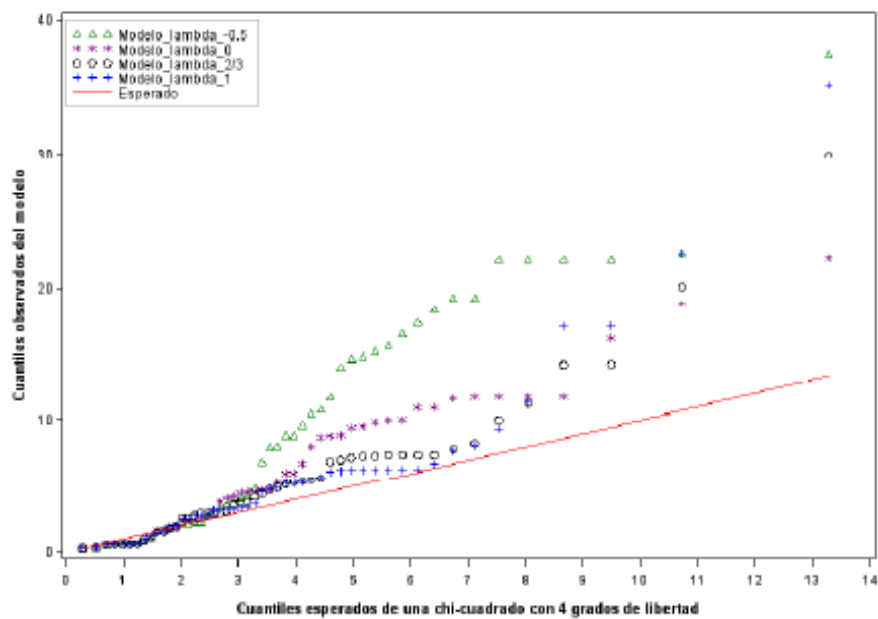


Figura 4.13. Gráfico cuantil-cuantil para la media de 100 cuantiles simulados para el Modelo 4

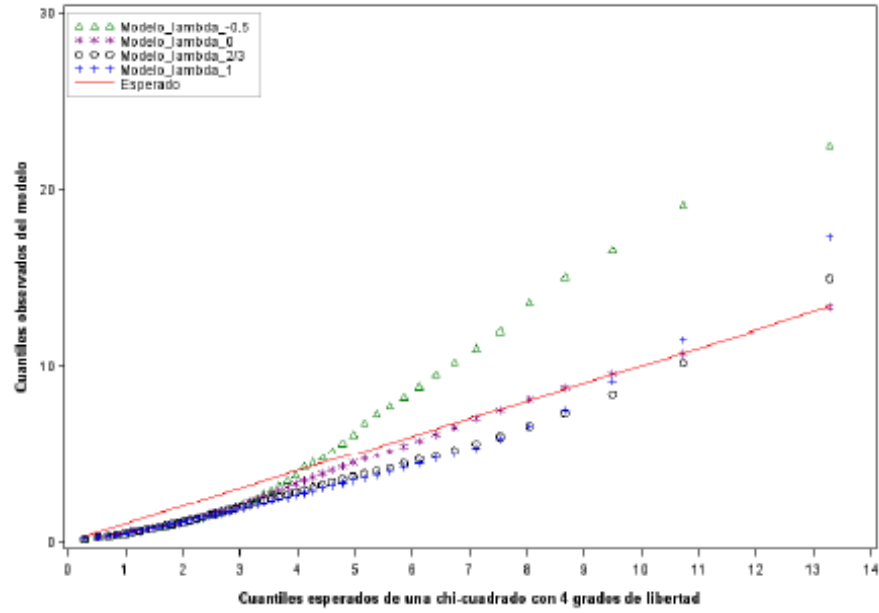


Figura 4.14. Gráfica de la banda min-max para 100 cuantiles simulados para el Modelo 1

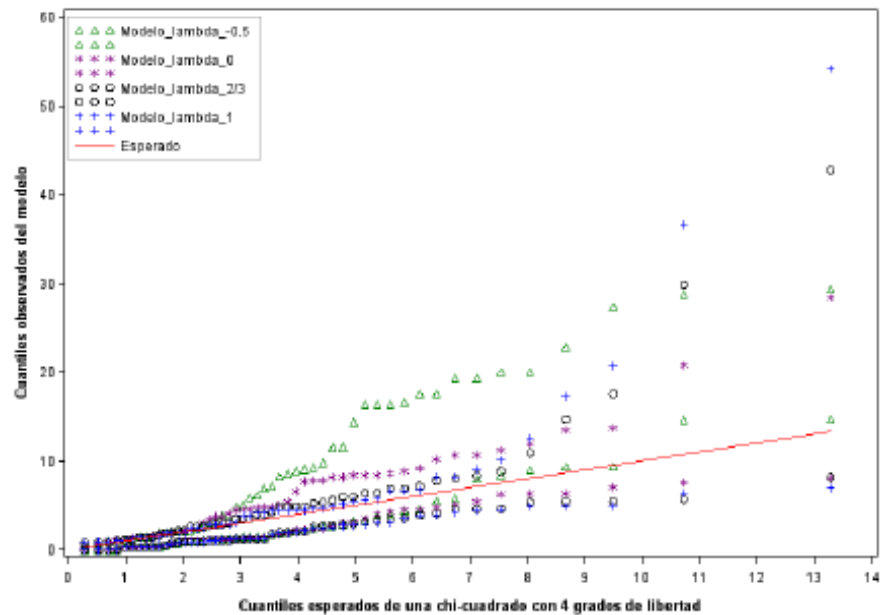




Figura 4.15. Gráfica de la banda min-max para 100 cuantiles simulados para el Modelo 2

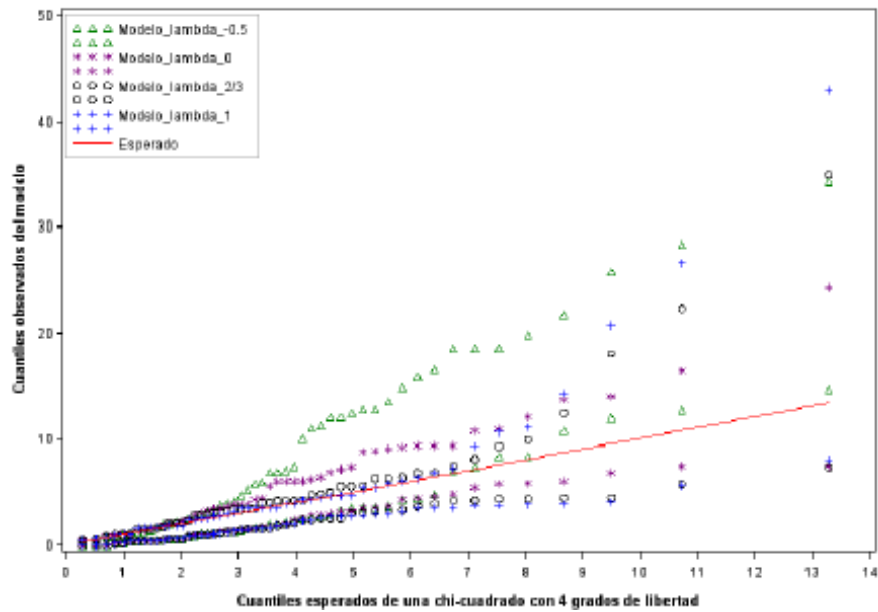


Figura 4.16. Gráfica de la banda min-max para 100 cuantiles simulados para el Modelo 3

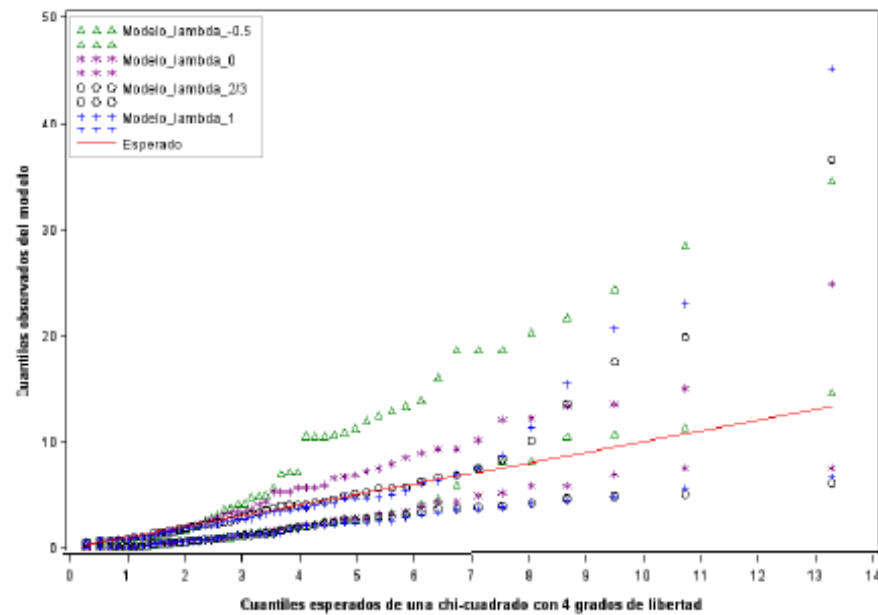
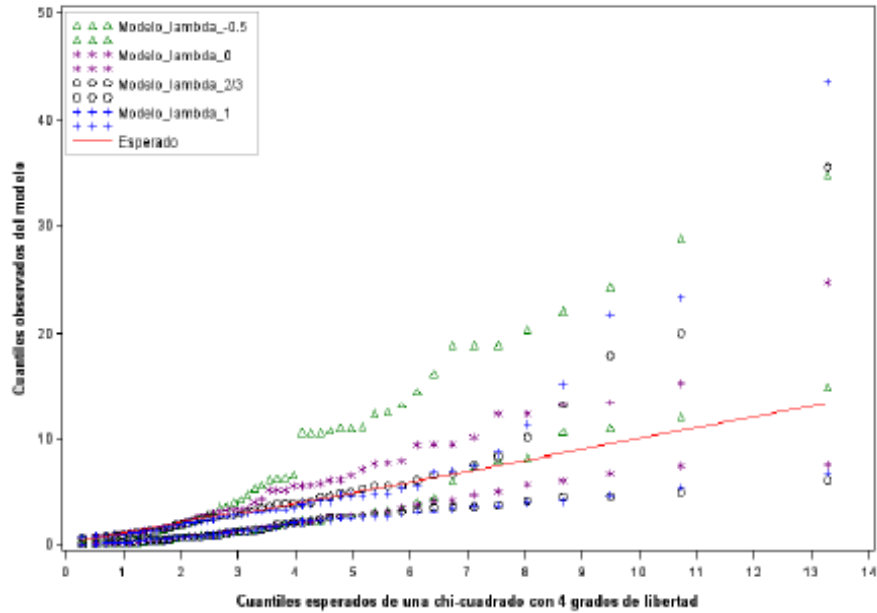


Figura 4.17. Gráfica de la banda min-max para 100 cuantiles simulados para el Modelo 4



#### 4.4. Medidas basadas en el volumen de elipsoides de confianza

En primer lugar, se define que se entiende por influencia. La definición, que nos parece más apropiada, fue dada por Besley, Kuh, y Welsch [18]:

*Una observación es influyente si individual o colectivamente tiene gran influencia sobre el cálculo de valores estimados comparadas con las otras observaciones del conjunto...este es el caso de la mayoría de las observaciones.*

Esta definición es subjetiva pero implica que podemos ordenar las observaciones de manera razonable de acuerdo a alguna medida de su influencia.

Una técnica para descubrir observaciones influyentes consiste en examinar el impacto que se produce al quitar todas las medidas de un individuo y anotar los resultados del efecto en los diferentes valores calculados. Los individuos que al ser quitadas sus medidas produzcan grandes cambios en el cálculo de valores asociados al modelo serán consi-

derados influyentes. Una observación, sin embargo, puede no tener la misma influencia sobre todos los valores calculados. La cuestión “en qué influye” es, por tanto, bastante importante. Una observación puede tener influencia sobre  $\hat{\beta}$ , los valores ajustados, y/o el estadístico de bondad de ajuste. El primer objetivo del análisis es determinar qué influencia debemos considerar. Por ejemplo, si el interés primordial es la estimación de  $\beta$ , entonces la medición de la influencia de las observaciones sobre  $\hat{\beta}$  es la indicada, mientras que si el primer objetivo es la predicción, entonces una medida de la influencia sobre los valores predichos es más apropiada que la medición de la influencia sobre  $\hat{\beta}$ .

Algunas de las medidas de influencia más conocidas, definidas en Preisser y Qaqish [91], que están implementadas en SAS son:

Un **punto leverage** es una observación grande en comparación con las otras observaciones. El “leverage” del individuo  $i$  está contenido en la matriz  $\mathbf{H}_i$  definida en (4.1). Una medida del “leverage” del individuo  $i$  puede ser definida como la traza de  $\mathbf{H}_i$

$$\text{Leverage}C_i = \text{traza}(\mathbf{H}_i) \quad (4.7)$$

La influencia del individuo  $i$  en los coeficientes de regresión puede ser valorada con la **distancia de Cook**

$$\text{Cook}C_i = \left( \hat{\beta}_{EEG} - \hat{\beta}_{EEG}^{(i)} \right)^T \mathbf{J}_1^{-1} \left( \hat{\beta}_{EEG} - \hat{\beta}_{EEG}^{(i)} \right) \frac{1}{p\hat{\gamma}} \quad (4.8)$$

donde  $\hat{\beta}_{EEG}^{(i)}$  denota el parámetro de regresión estimado cuando el individuo  $i$  es eliminado de los datos y  $p$  es el número de parámetros de regresión. Esta medida se puede aproximar por

$$\begin{aligned} \text{Cook}C_i = & (\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Lambda}_i^{-1} (\mathbf{W}_i^{-1} - \mathbf{X}_i \mathbf{J}_1^{-1} \mathbf{X}_i^T)^{-1} \mathbf{X}_i \mathbf{J}_1^{-1} \mathbf{X}_i^T \\ & \times (\mathbf{W}_i^{-1} - \mathbf{X}_i \mathbf{J}_1^{-1} \mathbf{X}_i^T)^{-1} \boldsymbol{\Lambda}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \frac{1}{p\hat{\gamma}} \end{aligned}$$

La **medida de influencia DFFITC<sub>i</sub>** de una observación es una medida de la influencia del individuo  $i$  sobre el ajuste global. Este nombre lo dieron Besley et al. [18] (difference between fitted values, es decir, distancia entre valores predichos),

$$\text{DFFITC}_i = (\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Lambda}_i^{-1} (\mathbf{W}_i^{-1} - \mathbf{X}_i \mathbf{J}_1 \mathbf{X}_i^T)^{-1} \mathbf{H}_i \boldsymbol{\Lambda}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \frac{1}{p\hat{\gamma}} \quad (4.9)$$

Por tanto, la medida (4.8) mide la influencia de un individuo en  $\hat{\beta}_{EEG}$  mientras que la medida (4.9) mide la influencia de un individuo en  $\hat{\beta}_{EEG}$  y la varianza estimada de  $\hat{\beta}_{EEG}$  simultáneamente.

Finalmente, se considera un estadístico estandarizado que mide el efecto de un único individuo sobre cada parámetro de regresión del modelo,

$$DFBETAC_{ij} = \frac{\mathbf{J}_1 \mathbf{X}_i^T (\mathbf{W}_i^{-1} - \mathbf{X}_i \mathbf{J}_1 \mathbf{X}_i^T)^{-1} \mathbf{\Lambda}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)}{\hat{\gamma} [\mathbf{J}_1]_{jj}^{1/2}} \quad (4.10)$$

Todas estas medidas son una generalización de herramientas de diagnóstico disponibles para MLG que a su vez son conocidas medidas de influencia en regresión lineal. Sin embargo, a nuestro conocimiento, las medidas basadas en el volumen de los elipsoides de confianza en regresión lineal no han sido generalizadas a EEG. Por ello, a continuación se propone una medida de la influencia de la  $i$ -ésima observación basada en el cambio del volumen de la elipsoide de confianza cuando la  $i$ -ésima observación es eliminada para EEG. Como sugieren Belsley et al. [18] en regresión lineal, la influencia de la  $i$ -ésima observación en la matriz de covarianza de los coeficientes de regresión estimados puede ser medido comparando la matriz de covarianza usando todos los datos con la matriz de covarianza que resulta cuando quitamos la  $i$ -ésima fila. Una de las maneras más simples de comparar dos matrices simétricas positivas es el cociente de sus determinantes. A continuación se define una extensión de esta medida para evaluar modelos de regresión con medidas repetidas:

$$CVR_i = \frac{Det(\mathbf{J}_1^{(i)} \mathbf{J}_2^{(i)} \mathbf{J}_1^{(i)})}{Det(\mathbf{J}_1 \mathbf{J}_2 \mathbf{J}_1)} \quad (4.11)$$

donde  $\mathbf{J}_1$  está definido en (4.2),

$$\mathbf{J}_2 = \sum_{i=1}^n \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^T [\mathbf{V}_i(\hat{\alpha})]^{-1} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i) (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)^T [\mathbf{V}_i(\hat{\alpha})]^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}.$$

y el superíndice  $(i)$  indica que se ha eliminado la fila  $i$  (individuo  $i$ ).

Los valores de esta medida próximos a 1 indican que las dos matrices de covarianza no se ven afectadas por la eliminación del individuo  $i$ . Pero observaciones cuyos valores resultantes no están próximos a 1, son posiblemente influyentes y es necesario un estudio

más detallado. Por eso, se define la medida  $C_i$  como

$$C_i = |CVR_i - 1|. \quad (4.12)$$

Es decir, de esta medida estudiaremos los puntos con valores grandes.

Teniendo en cuenta el Teorema 3.2, sabemos que la función de densidad de  $\hat{\beta}_{EEG}$ ,  $f_{\hat{\beta}_{EEG}}(x)$ , es aproximadamente una normal  $p$ -dimensional con vector de medias  $\beta$  y matriz de covarianzas  $\mathbf{V}_{\hat{\beta}_{EEG}}$ . De forma análoga podemos probar que la función de densidad de  $\hat{\beta}_{EEG}^{(i)}$ ,  $f_{\hat{\beta}_{EEG}^{(i)}}(x)$ , es aproximadamente una normal  $p$ -dimensional con vector de medias  $\beta$  y matriz de covarianzas  $\mathbf{V}_{\hat{\beta}_{EEG}^{(i)}}$  donde  $\hat{\beta}_{EEG}^{(i)}$  es el parámetro de regresión estimado cuando el individuo  $i$  ha sido eliminado de los datos.

Nuestra propuesta es medir la discrepancia entre las funciones de densidad asintóticas  $f_{\hat{\beta}_{EEG}}(x)$  y  $f_{\hat{\beta}_{EEG}^{(i)}}(x)$  para comparar la matriz de covarianzas obtenida con todos los datos, con la matriz de covarianzas que resulta cuando quitamos el individuo  $i$ . Utilizamos la medida de divergencia de Kullback- Leibler definida por

$$D_{KULLBACK}(\hat{\beta}_{EEG}, \hat{\beta}_{EEG}^{(i)}) = \int_{\mathbb{R}^p} f_{\hat{\beta}_{EEG}}(x) \log \left( \frac{f_{\hat{\beta}_{EEG}}(x)}{f_{\hat{\beta}_{EEG}^{(i)}}(x)} \right) dx,$$

llamada Información Mínima discriminante entre las funciones de densidad de  $f_{\hat{\beta}_{EEG}}(x)$  y  $f_{\hat{\beta}_{EEG}^{(i)}}(x)$ , para medir esta discrepancia (ver Kullback y Leibler [57]). Teniendo en cuenta la expresión de la divergencia de Kullback entre dos distribuciones normales multivariantes que se puede ver en Pardo [79], se propone una segunda medida de la influencia del individuo  $i$ -ésimo en la matriz de covarianza de los coeficientes de regresión estimados dada por

$$DCV_i = \frac{1}{2} \text{traza} \left\{ \left[ \left( \mathbf{J}_1^{(i)} \mathbf{J}_2^{(i)} \mathbf{J}_1^{(i)} \right)^{-1} (\mathbf{J}_1 \mathbf{J}_2 \mathbf{J}_1) - \mathbf{I} \right] + \log \frac{\text{Det}(\mathbf{J}_1^{(i)} \mathbf{J}_2^{(i)} \mathbf{J}_1^{(i)})}{\text{Det}(\mathbf{J}_1 \mathbf{J}_2 \mathbf{J}_1)} \right\}. \quad (4.13)$$

Por supuesto, la generalización de esta medida considerando en lugar de la divergencia de Kullback cualquier otra divergencia es clara.

## 4.5. Ejemplos

### 4.5.1. Ejemplo respuesta Poisson

Para ilustrar las herramientas de diagnóstico introducidas en este capítulo se analizan los datos procedentes de un análisis clínico con 59 enfermos de epilepsia estudiados por Leppik et al. [60]. Los datos se muestran en la Tabla 4.1. Los pacientes que sufren ataques epilépticos simples o complejos son sometidos de forma aleatoria a un tratamiento activo antiepiléptico o a un placebo, como complemento a la medicación estándar. Se anota el número de ataques que sufren los pacientes tras serles asignados sus tratamientos en cada una de las cuatro visitas que se llevan a cabo en ocho semanas. Aunque cada paciente posteriormente se le administró el otro tratamiento, nosotros consideraremos sólo las cuatro respuestas primeras.

Por  $y_{ij}$  denotamos las respuestas (ataques) para el sujeto  $i$ -ésimo ( $i = 1, \dots, 59$ ) en el tiempo  $j$  ( $j = 2, 4, 6, 8$ ). Las variables en los modelos son: tasa de ataques (Base, logaritmo de un cuarto del n° de ataques computados en las 8 semanas del tratamiento), el logaritmo de edad (Edad), el indicador binario para el grupo de progabide, Trt (0 placebo, 1 tratamiento activo), y las visitas clínicas (Visit).

Thall y Vail [108] consideraron el modelo:

$$\log(\mu_{ij}) = \beta_0 + \beta_1 \cdot Trt + \beta_2 \cdot Edad + \beta_3 \cdot Base + \beta_4 \cdot Base \cdot Trt + \beta_5 \cdot Visit \quad (4.14)$$

que ajustamos utilizando la metodología EEG con una matriz de trabajo con estructura de correlación intercambiable.

Comparamos las nuevas medidas de influencia introducidas en este capítulo con algunos diagnósticos de regresión para el análisis de las ecuaciones de estimación generalizadas (EEG) disponibles en el procedimiento estándar de SAS PROC GENMOD en su versión 9.2 y descritas en la Sección 4.4.

En las Figuras 4.18-4.21 se presentan los residuos de Pearson,  $c_{it}^{\phi(1)}$ , de desviación,  $c_{it}^{\phi(0)}$  y  $c_{it}^{\phi(2/3)}$  descritos en (3.5) para los 4 instantes de tiempo. Los residuos más grandes para  $t = 1$  son para la observaciones 10 y 35 para cualquier valor de  $\lambda$ . El mayor residuo en  $t = 2$  es para la observación 56, mientras que en  $t = 3$  el mayor residuo lo tiene la observación 25. Es decir, éstas son las observaciones peor predichas en cada instante de

tiempo.

Tabla 4.1

ID	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Trt	Base	Edad	ID	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Trt	Base	Edad
104	5	3	3	3	0	11	31	103	0	4	3	0	1	19	20
106	3	5	3	3	0	11	30	108	3	6	1	3	1	10	30
107	2	4	0	5	0	6	25	110	2	6	7	4	1	19	18
114	4	4	1	4	0	8	36	111	4	3	1	3	1	24	24
116	7	18	9	21	0	66	22	112	22	17	19	16	1	31	30
118	5	2	8	7	0	27	29	113	5	4	7	4	1	14	35
123	6	4	0	2	0	12	31	117	2	4	0	4	1	11	27
126	40	20	23	12	0	52	42	121	3	7	7	7	1	67	20
130	5	6	6	5	0	23	37	122	4	18	2	5	1	41	22
135	14	13	6	0	0	10	28	124	2	1	1	0	1	7	28
141	26	12	6	22	0	52	36	128	0	2	4	0	1	22	23
145	12	6	8	4	0	33	24	129	5	4	0	3	1	13	40
201	4	4	6	2	0	18	23	137	11	14	25	15	1	46	33
202	7	9	12	14	0	42	36	139	10	5	3	8	1	36	21
205	16	24	10	9	0	87	26	143	19	7	6	7	1	38	35
206	11	0	0	5	0	50	26	147	1	1	2	3	1	7	25
210	0	0	3	3	0	18	28	203	6	10	8	8	1	36	26
213	37	29	28	29	0	111	31	204	2	1	0	0	1	11	25
215	3	5	2	5	0	18	32	207	102	65	72	63	1	151	22
217	3	0	6	7	0	20	21	208	4	3	2	4	1	22	32
219	3	4	3	4	0	12	29	209	8	6	5	7	1	41	25
220	3	4	3	4	0	9	21	211	1	3	1	5	1	32	35
222	2	3	3	5	0	17	32	214	18	11	28	13	1	56	21
226	8	12	2	8	0	28	25	218	6	3	4	0	1	24	41
227	18	24	76	25	0	55	30	221	3	5	4	3	1	16	32
230	2	1	2	1	0	9	40	225	1	23	19	8	1	22	26
234	3	1	4	2	0	10	19	228	2	3	0	1	1	25	21
238	13	15	13	12	0	47	22	232	0	0	0	0	1	13	36
101	11	14	9	8	1	76	18	236	1	4	3	2	1	12	37
102	8	7	9	4	1	38	32								

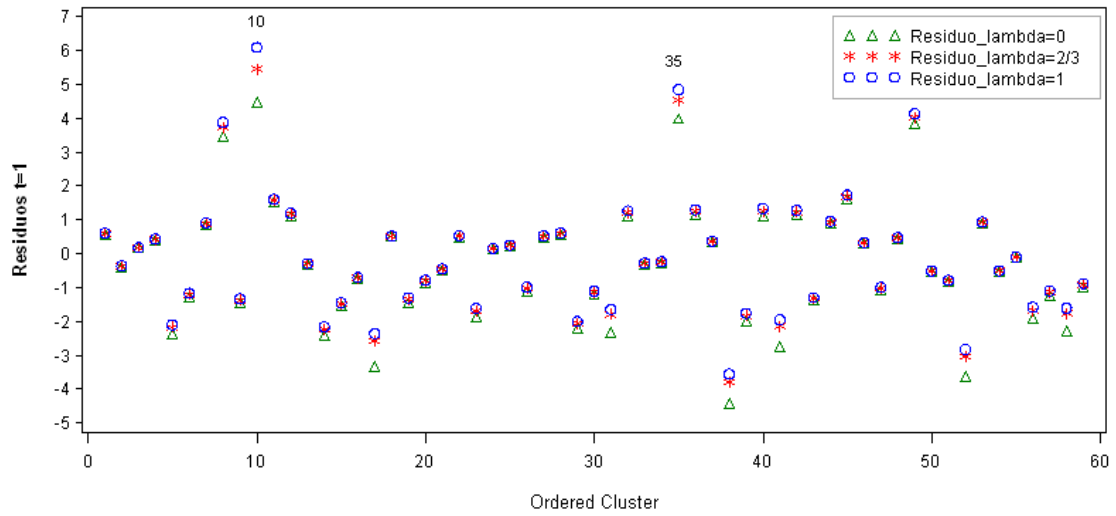
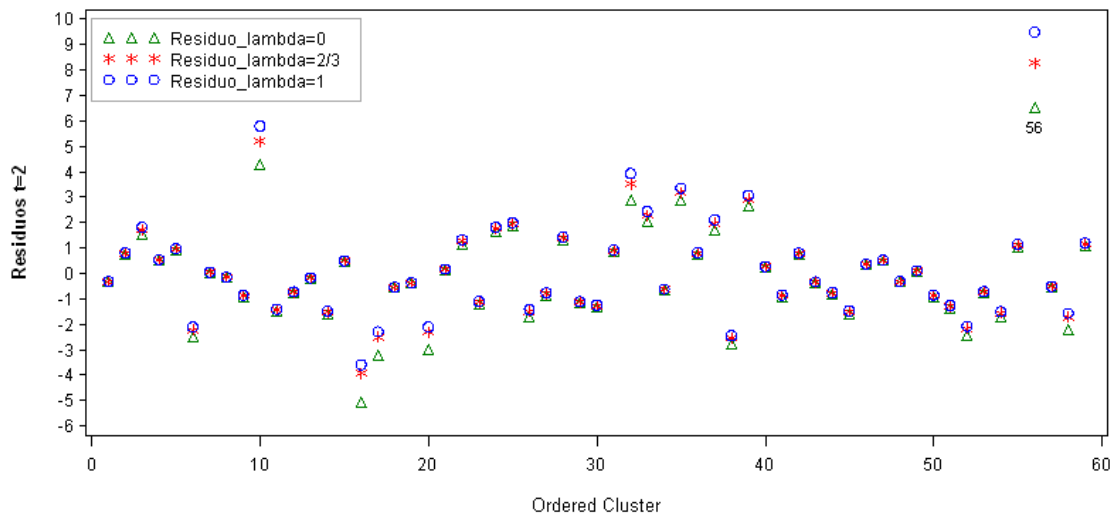
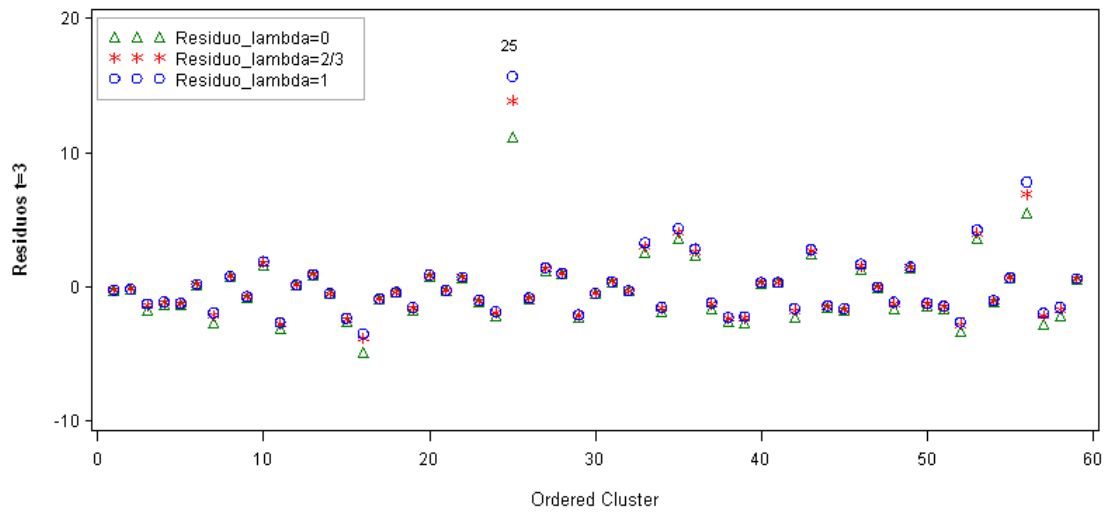
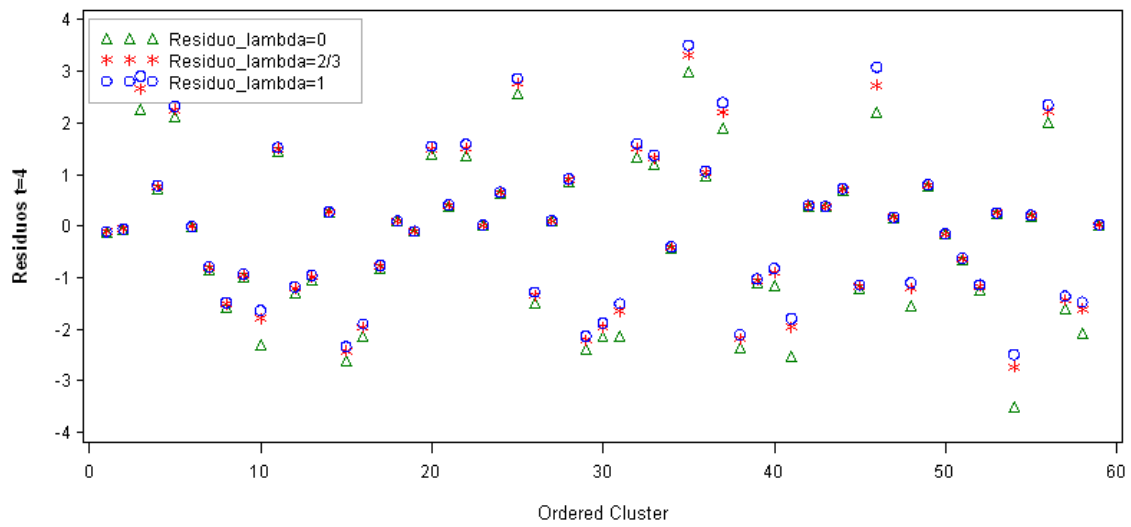
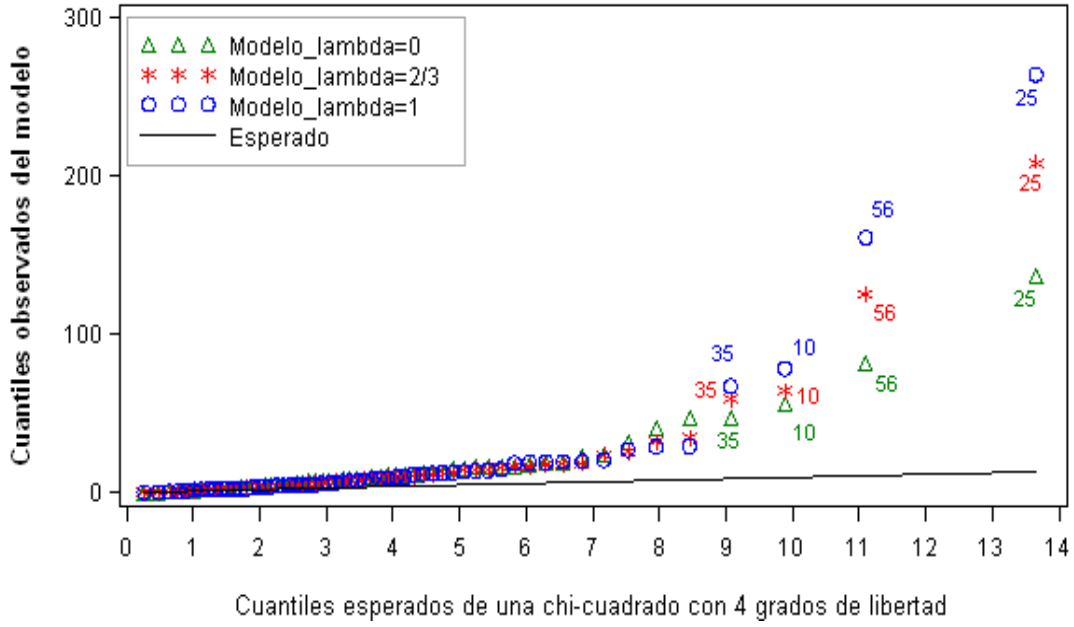
Figura 4.18. Gráfico de los residuos  $c_{i1}^{\phi(\lambda)}$ Figura 4.19. Gráfico de los residuos  $c_{i2}^{\phi(\lambda)}$ 



Figura 4.20. Gráfico de los residuos  $c_{i3}^{\phi(\lambda)}$ Figura 4.21. Gráfico de los residuos  $c_{i4}^{\phi(\lambda)}$ 

En la Figura 4.22 se presenta el gráfico cuantil-cuantil descrito en la Sección 4.2 para los datos del ejemplo. Se puede ver que las observaciones 25, 56, 10 y 35 que eran las que aparecen como “outlying” en algún tiempo  $t$  al graficar los residuos  $c_{it}^{\phi(\lambda)}$  siguen siendo las que este gráfico apunta como “outlying” también.

Figura 4.22. Gráfico cuantil-cuantil para los datos de la Tabla 4.1

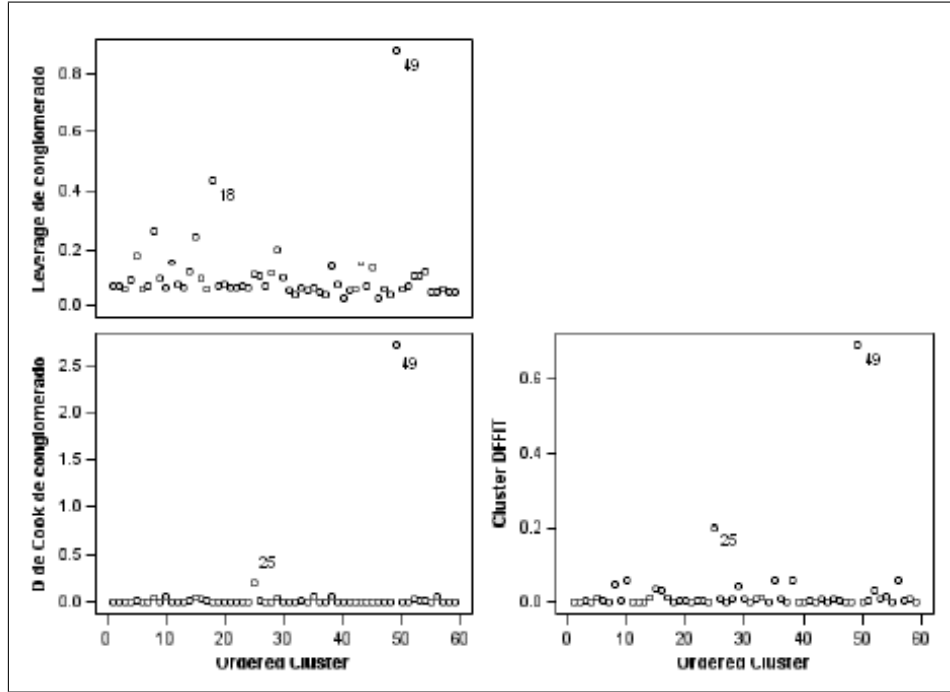


La medida que calcula SAS cuando se utiliza el procedimiento EEG para detectar puntos “leverage” es  $leverageC_i$  definida en (4.7) y para detectar observaciones influyentes son  $CookC_i$  y  $DFFITC_i$  definidas en (4.8) y (4.9), respectivamente. En la Figura 4.23-4.25, aparecen representadas dichas medidas.

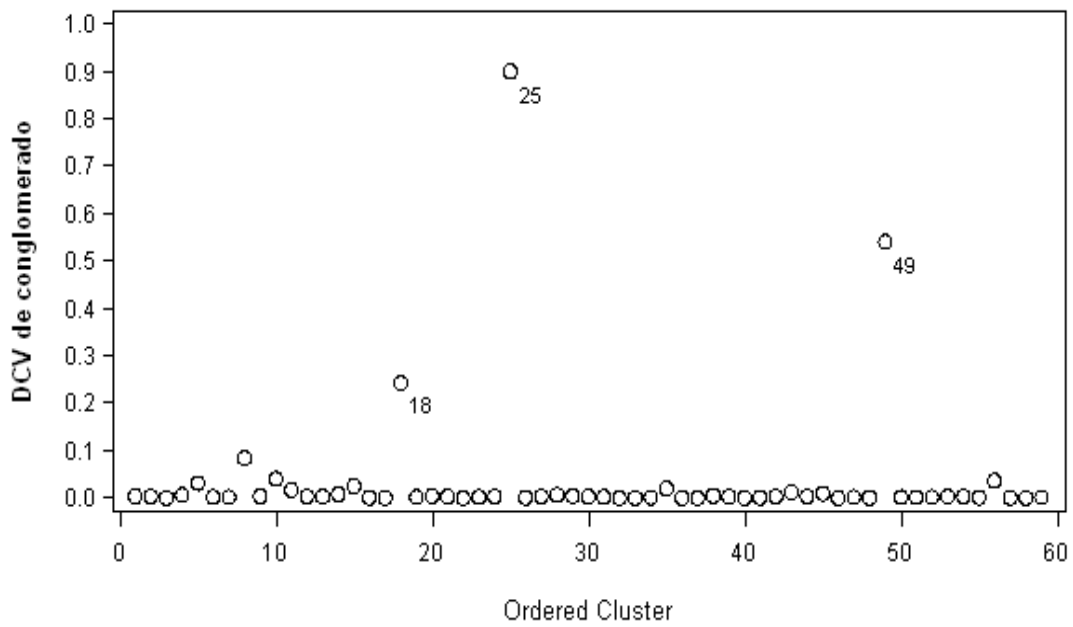
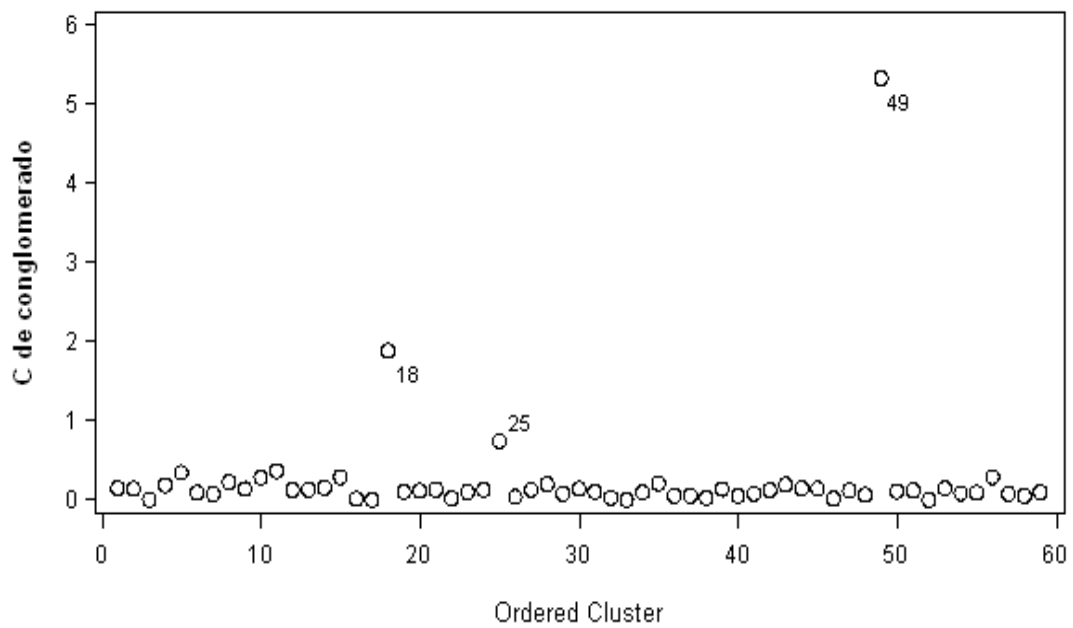
Figura 4.23 (arriba). Medida  $\text{LeverageC}_i$

Figura 4.24 (abajo e izquierda). Distancia de Cook,  $\text{CookC}_i$

Figura 4.25 (abajo y derecha). Medida influencia  $\text{DFFITC}_i$ .



La Figura 4.23 muestra como puntos “leverage” las observaciones 49 y 18. Las Figuras 4.24 y 4.25 representan la distancia de *Cook* definida en (4.8) y *DFFITC* definida en (4.9). Ambas medidas apuntan las observaciones 49 y 25 como posibles datos influyentes para los parámetros. Una de estas dos observaciones, la 49, apareció también como un punto “leverage”. La Figura 4.26 muestra la nueva medida  $\text{DCV}_i$  definida en (4.12). Las observaciones 25, 49 y 18 apuntan como observaciones influyentes para esta medida. Finalmente, la Figura 4.27 muestra la medida  $C_i$  definida en (4.13). Esta medida detecta las mismas observaciones influyentes que  $\text{DCV}_i$  aunque no en el mismo orden. Para esta medida las observaciones más influyentes en orden decreciente son 49, 18 y 25.

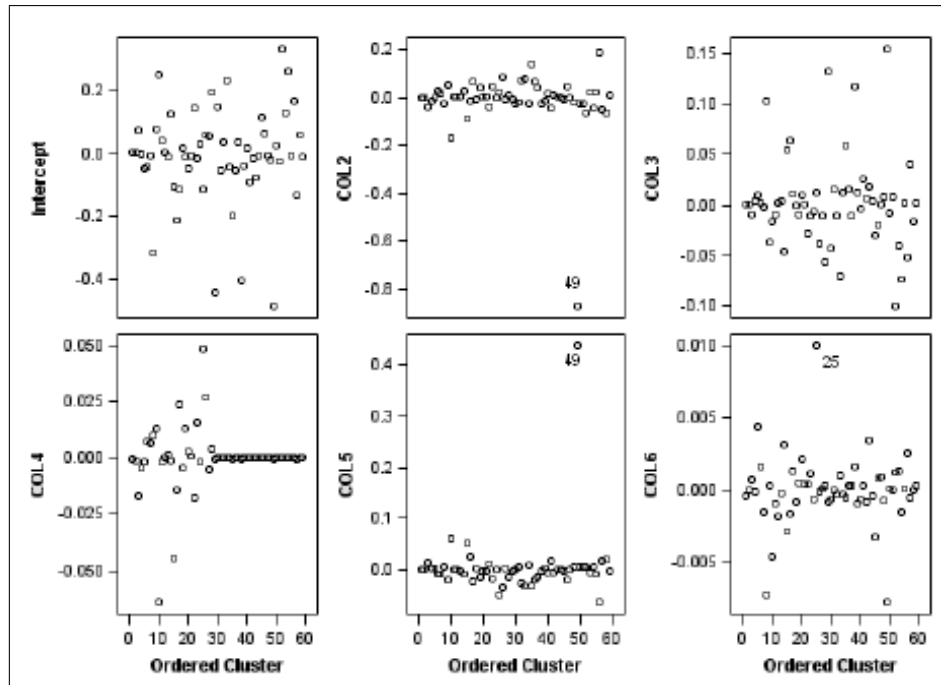
Figura 4.26. Medida de Influencia  $DCV_i$ Figura 4.27. Medida de Influencia  $C_i$ 

En la Tabla 4.2 aparece un resumen de las medidas consideradas y las observaciones que toman un valor grande de cada una de ellas en orden decreciente.

Tabla 4.2

Medidas basadas en	Medidas influencia	Observaciones Influyentes
Gráfico cuantil-cuantil	$q_i^\phi$	25, 56, 10, 34
Matriz hat	$leverage C_i$	49, 18
Volumen del elipsoide de confianza	$C_i$ $DCV_i$	49, 18, 25 25, 49, 18
Función de Influencia	$Cook C_i$ $DFFIT C_i$	49, 25 49, 25

Figura 4.28 Medida DFBETA para cada variable



Col2=Trt, Col3=Edad, Col4=Base, Col5=Base  $\times$  Trt, Col6=Visit

Las conclusiones de las medidas basadas en la función de influencia coinciden, mientras que las basadas en el volumen del elipsoide de confianza no. La observación 49 aparece como la más influyente para las medidas  $leverageC_i$ ,  $C_i$ ,  $CookC_i$  y  $DFFITC_i$ , mientras que el gráfico cuantil-cuantil y  $DCV_i$  obtienen la observación 25 como la más influyente.

Por otra parte, las medidas  $C_i$  y  $DCV_i$  obtienen más puntos influyentes que el resto. Esto nos podría hacer pensar que son medidas conservadoras. Sin embargo, tal y como se puede ver en la Figura 4.28 en las que se representan las medidas  $DFBETAC_i$ ; definidas por (4.10) cada una de estas observaciones es influyente en al menos una variable.

Si echamos un vistazo a los datos, estos nos sugieren que la observación 49 es un “outlyrer”, ya que tanto los valores de la variable Base como la del postratamiento son mucho más grandes que los de las demás observaciones. Sin embargo, se ha visto que las medidas diagnóstico no apuntan a esta observación como un “outlying” sino como una observación influyente. Para estudiar la influencia que tiene esta observación sobre las estimaciones de los parámetros, en la Tabla 4.3 se presentan las estimaciones de los parámetros del modelo y sus errores estándar considerando todos los datos y quitando la observación 49.

Tabla 4.3. Los estimadores de (4.14) y sus errores estándar			
(a) Todos los datos		(b) Quitando la observación 49	
$\hat{\beta}$	E.E	$\hat{\beta}$	E.E
-2.7131	0.9453	-2.2950	0.8734
-1.3571	0.4283	-0.5258	0.4150
0.9161	0.2758	0.7807	0.2547
0.9480	0.0976	0.9492	0.0972
0.5708	0.1740	0.1404	0.1928
-0.0294	0.0175	-0.0214	0.0189

La estimación para el parámetro de la variable Visit disminuye en 2.47 veces el error estándar, mientras que la estimación del parámetro de la variable Edad disminuye en 1.94 veces el error estándar.

A continuación, se repite el estudio eliminando la observación 49 del conjunto de datos. En las Figuras 4.29-4.32 se presentan las gráficas de los residuos  $c_{it}^{\phi(\lambda)}$  con  $\lambda = 0, 2/3$  y 1 para  $t = 1, 2, 3$  y 4, respectivamente. Estas gráficas detectan las mismas observaciones que antes como “outlying”.

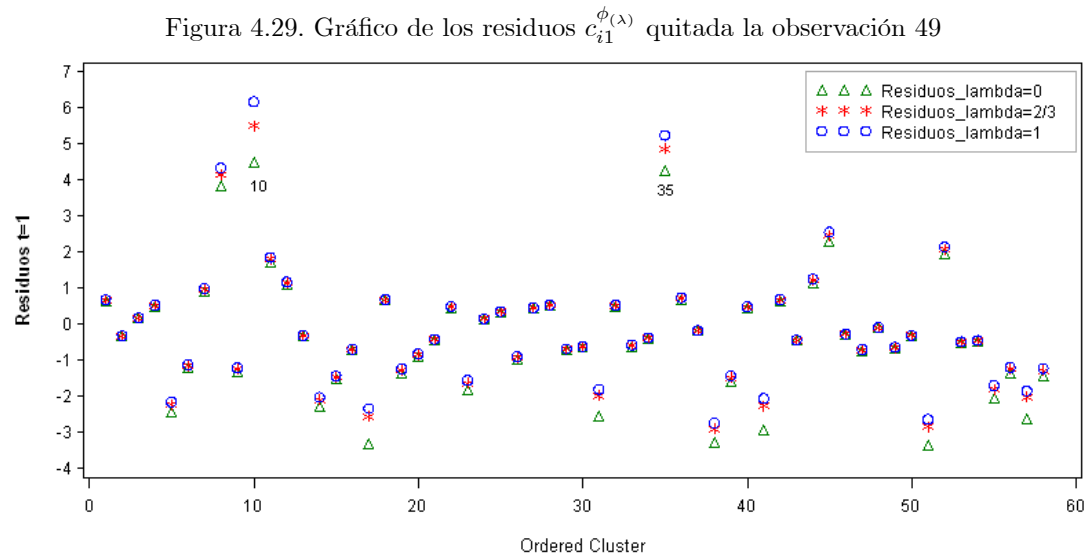


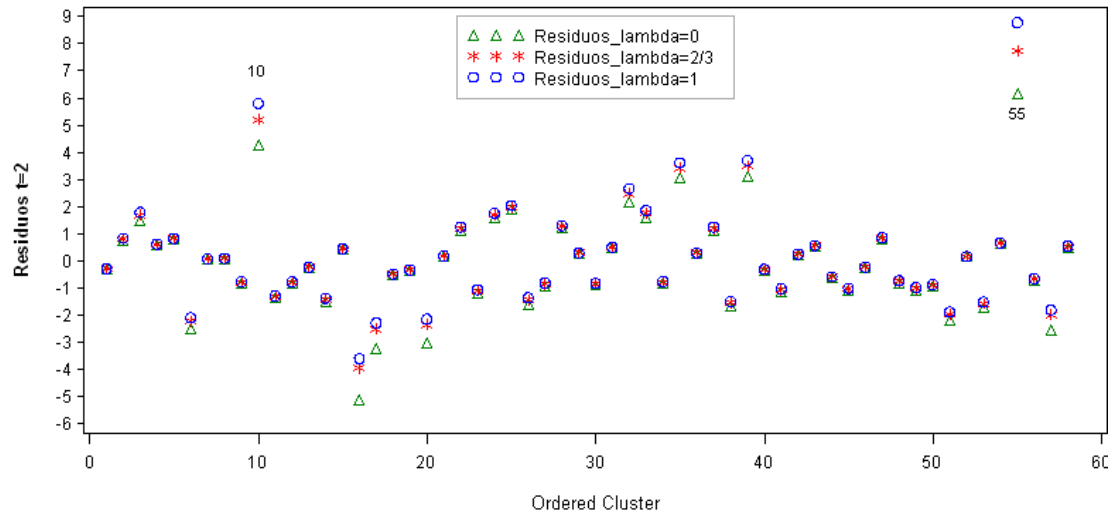
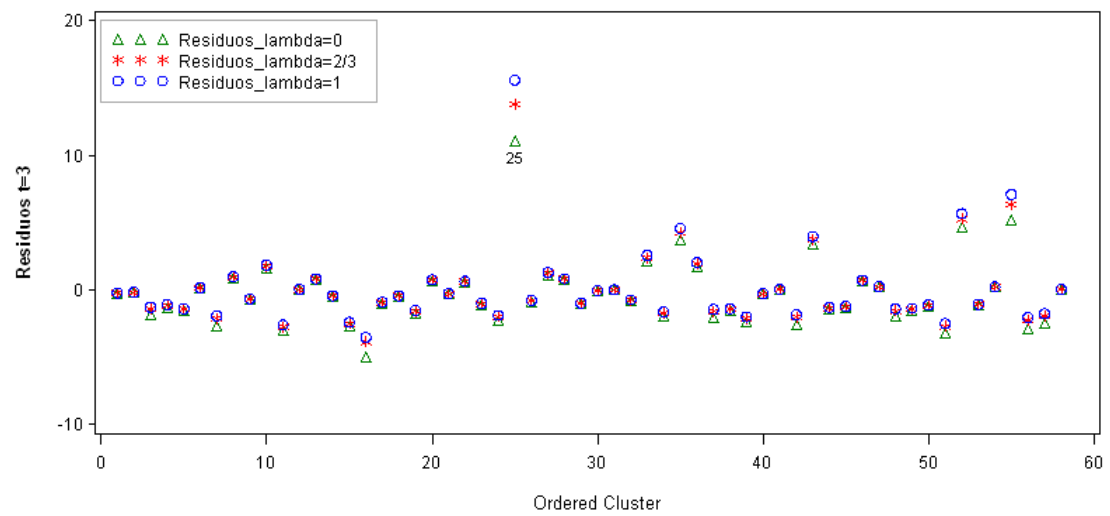
Figura 4.30. Gráfico de los residuos  $c_{i2}^{\phi(\lambda)}$  quitada la observación 49Figura 4.31. Gráfico de los residuos  $c_{i3}^{\phi(\lambda)}$  quitada la observación 49



Figura 4.32. Gráfico de los residuos  $c_{i4}^{\phi(\lambda)}$  quitada la observación 49

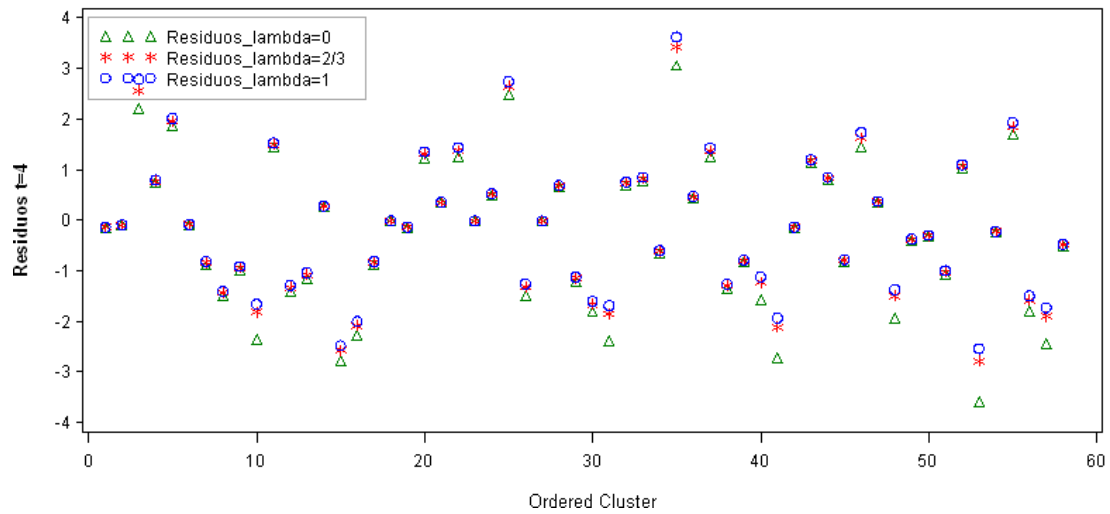
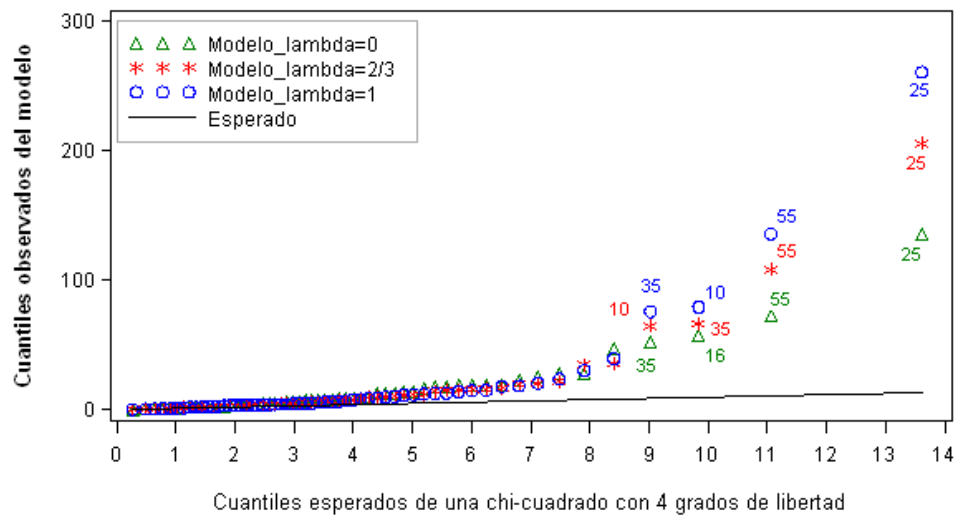


Figura 4.33. Gráfico cuantil-cuantil para los datos de la Tabla 4.1 quitada la observación 49.



Las Figuras 4.34-4.36 son obtenidas ajustando el modelo al conjunto de datos inicial al que le hemos quitado la fila 49. La Figura 4.34 muestra como punto de “leverage” de nuevo la observación 18. Las Figuras 4.35 y 4.36 representan la distancia de  $CookC_i$  y  $DFFITC_i$  definidas en (4.8) y en (4.9), respectivamente. La observación 25 aparece como un posible dato influyente. La Figura 4.37 muestra que las observaciones 25 y 18 tienen el mayor valor de  $DCV_i$  definido en (4.13). Finalmente, La Figura 4.38 muestra la medida  $C_i$  definida en (4.12) que detecta como influyente la observación 18 que aparece como punto “leverage” en la Figura 4.34.

Figura 4.34 (arriba). Medida  $LeverageC_i$

Figura 4.35 (abajo e izquierda). Distancia de Cook,  $CookC_i$

Figura 4.36 (abajo y derecha). Medida influencia  $DFFITC_i$

en todas ellas hemos eliminado la observación 49

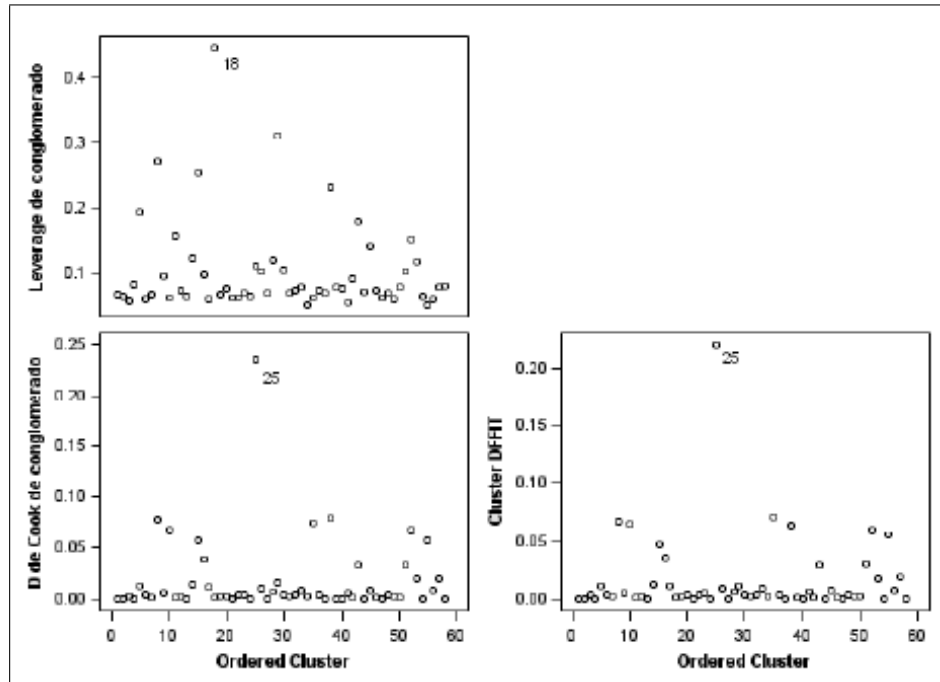
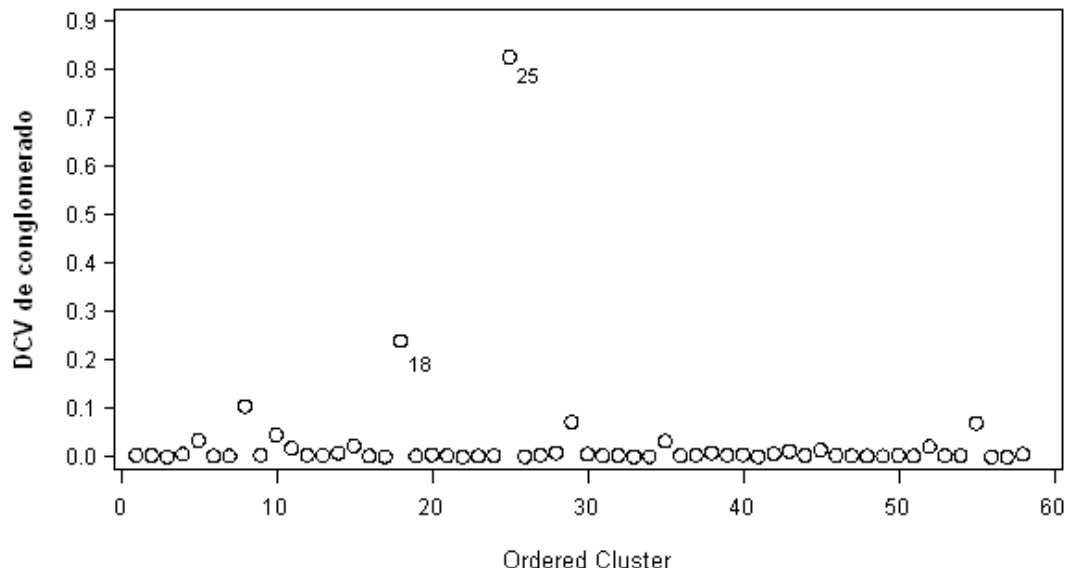
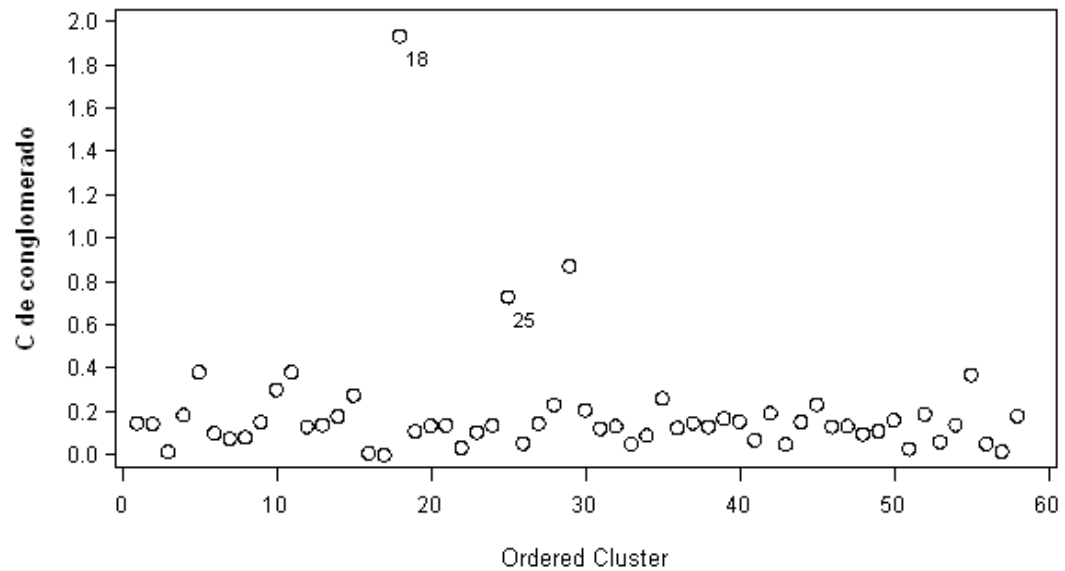
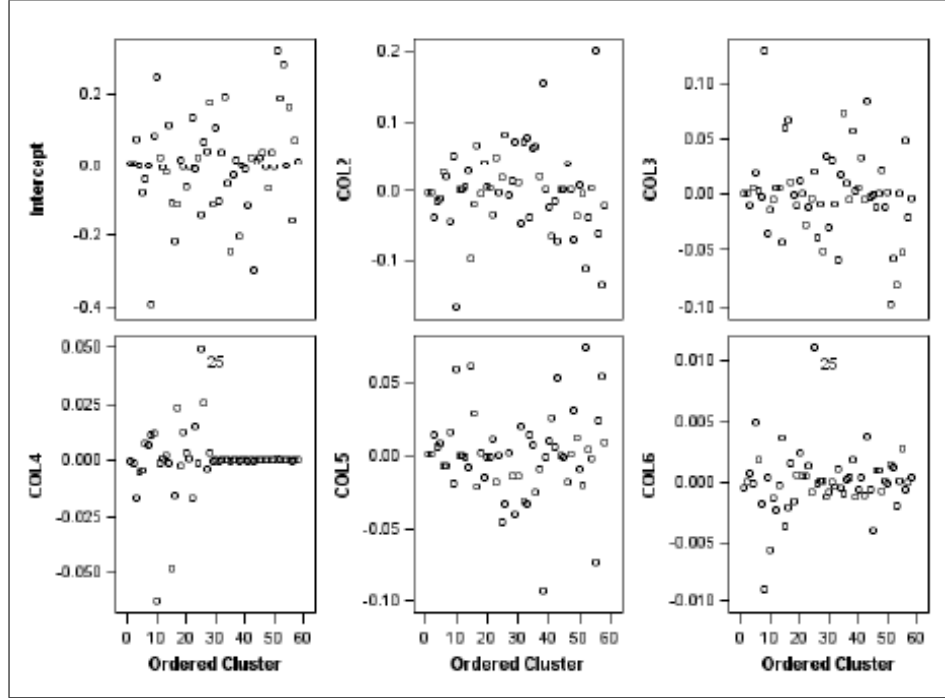


Figura 4.37 Medida de Influencia  $DCV_i$  quitada la observación 49Figura 4.38. Medida de Influencia  $C_i$  quitada la observación 49

La Figura 4.39 con la medida  $DFBETA$  no presentan puntos muy lejos del 0. Aunque la observación 25 toma valores más altos en las variables Base y Visit.

Figura 4.39 Medida  $DFBETA$  para cada variable quitada la observación 49



Col2=Trt, Col3=Edad, Col4=Base, Col5=Base  $\times$  Trt, Col6=Visit

Como conclusión, las nuevas medidas de diagnóstico detectan más observaciones influyentes que las medidas propuestas por Preisser y Qaqish [91] y en concreto la medida  $C_i$  detecta siempre con fuerza las observaciones “leverage” detectadas por la medida  $leverageC_i$  aparte de las demás observaciones influyentes. Otra cosa ocurre con la medida  $DCV_i$  detecta las observaciones leverage e influyentes y muestra la más influyente seguido de la de mayor leverage. Es decir,  $DCV_i$  y  $C_i$  parecen ser suficientes para detectar observaciones influyentes. Conclusiones similares se obtienen con la medida similar a  $C_i$  en regresión lineal.

### 4.5.2. Ejemplo respuesta binaria

Koch et al. [55] presentaron un estudio con datos longitudinales comparando un nuevo medicamento (N. medic) con un medicamento estándar (estándar) en el tratamiento de la depresión que sufrían 340 pacientes. Los pacientes fueron diagnosticados como enfermos con depresión “leve” o “grave”. En cada grupo, a los pacientes se les asignó uno de los dos medicamentos de forma aleatoria. Se les hizo un seguimiento las semanas 1, 2 y 4 de tratamiento. Cada paciente fue clasificado como normal o anormal en términos de padecimiento de la depresión. En este estudio, utilizamos las siguientes variables explicativas: diagnóstico (Diag), tratamiento (Trt) y tiempo (Tiempo) ( $j = 1, 2, 4$ ). Koch et al. [55] utilizaron los tiempos (0, 1, 2), que son el logaritmo en base 2 del número de semanas (1, 2, 4). Este conjunto de datos fue también analizado por Barnhart y Williamson [15], Agresti [2] y Oh et al. [76] y se presentan en la Tabla 4.4:

Tabla 4.4. Tabla de contingencia de respuestas en 3 tiempos  
(N=normal, A=anormal) para diagnóstico y tratamiento

Diag	Trt.	Respuestas a las semanas 1, 2 y 4								Total
		NNN	NNA	NAN	NAA	ANN	ANA	AAN	AAA	
Leve	Estándar	16	13	9	3	14	4	15	6	80
Leve	N. medic	31	0	6	0	22	2	9	0	70
Grave	Estándar	2	2	8	9	9	15	27	28	100
Grave	N. medic	7	2	5	2	31	5	32	6	90

Se considera el modelo propuesto por Oh et al. [76]

$$\log it(\mu_{ij}) = \beta_0 + \beta_1 \cdot Trt + \beta_2 \cdot Diag + \beta_3 \cdot Tiempo.$$

En la Tabla 4.5 se presenta un análisis de los estimadores del modelo obtenidos mediante el procedimiento EEG utilizando la matriz de correlación de trabajo intercambiable.

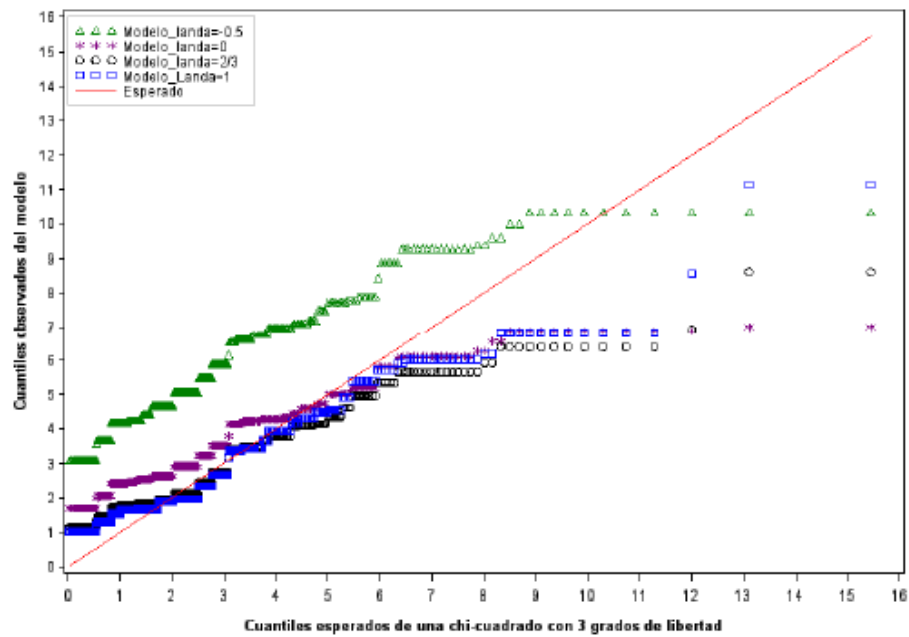
Tabla 4.5. Ajuste del modelo para el conjunto de datos Depresión.

Análisis de estimadores del parámetro EEG						
Estimadores del error estándar empírico						
Parámetro	Estimador	Error Estándar	95 % Límites de confianza		Z	Pr> Z
Intercepto	-0.5284	0.1395	-0.8013	-0.2555	-3.80	0.0001
Trt	0.9443	0.1383	0.6732	1.2155	6.83	<0.0001
Diag	-1.4675	0.1399	-1.7417	-1.1934	-10.49	<0.0001
Tiempo	0.9091	0.0938	0.7252	1.0930	9.69	<0.0001

En primer lugar, en la Figura 4.40, se presenta el gráfico cuantil-cuantil para los datos. Aunque los residuos  $q_i^{\phi(\lambda)}$  para  $\lambda = 0$  y  $2/3$  se separan más de la recta  $y = x$  que  $q_i^{\phi(1)}$  todos ellos muestran que este modelo es razonable.

Figura 4.40.

Gráfico cuantil-cuantil para los distintos valores de la Tabla 4.4



Además, tanto en los gráficos de las medidas de influencia disponibles en SAS, Figuras 4.41-4.43 como en los de las nuevas medidas introducidas en la Sección 4.4, presentadas en las Figuras 4.44-4.45 no destaca ningún punto influyente en el modelo.

Figura 4.41 (arriba). Medidas  $\text{LeverageC}_i$

Figura 4.42 (abajo e izquierda). Distancia de Cook,  $\text{CookC}_i$

Figura 4.43 (abajo y derecha). Medida Influencia  $\text{DFFITC}_i$

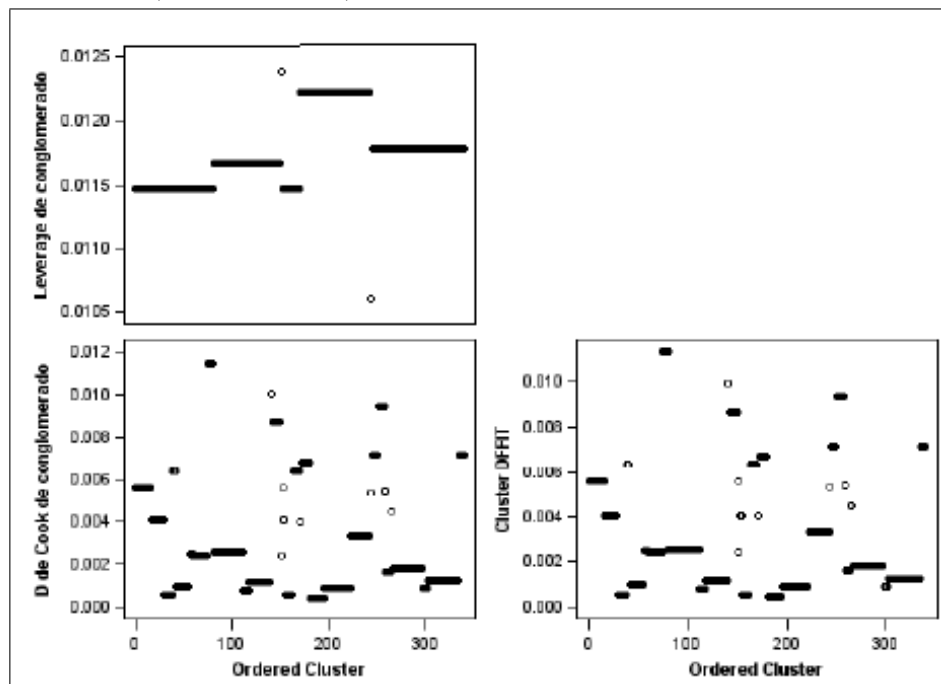
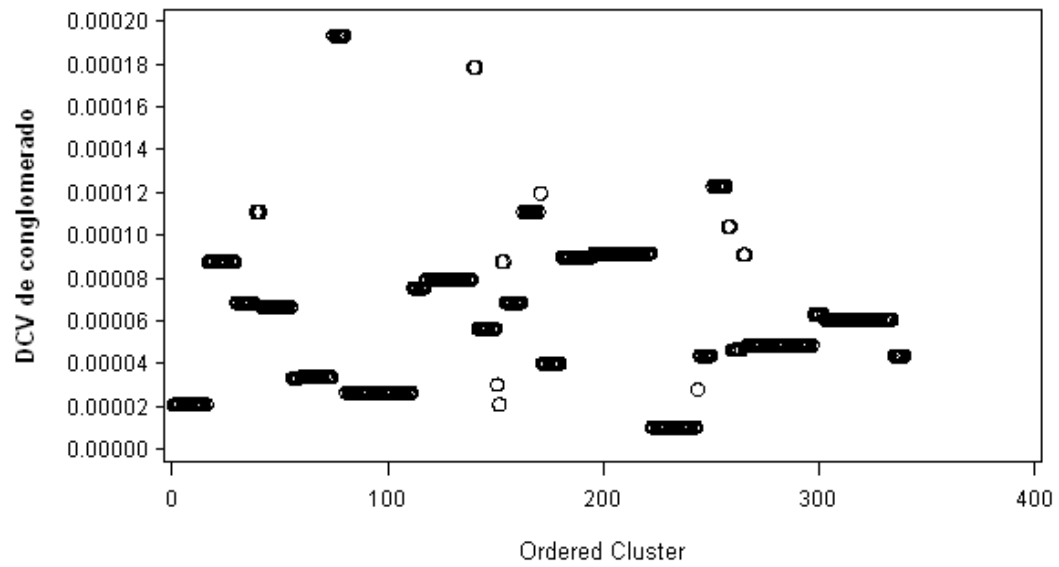
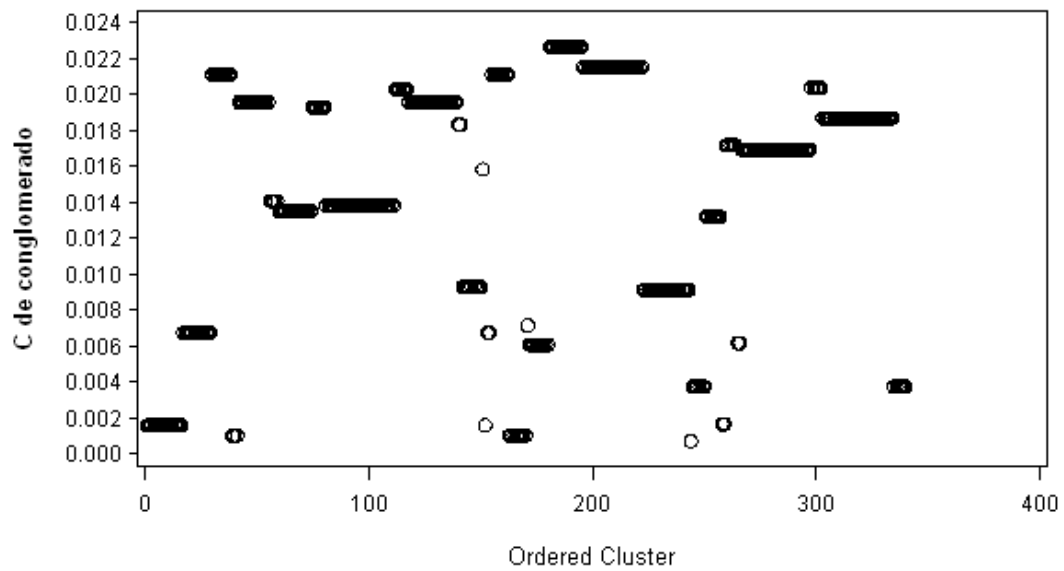


Figura 4.44. Medida de Influencia  $DCV_i$ Figura 4.45. Medida de Influencia  $C_i$ 



Si se ajustan a los datos alternativamente los modelos:

$$\text{Modelo 1: } \log it(\mu_{ij}) = \beta_0$$

$$\text{Modelo 2: } \log it(\mu_{ij}) = \beta_0 + \beta_1 \cdot Trt$$

$$\text{Modelo 3: } \log it(\mu_{ij}) = \beta_0 + \beta_1 \cdot Trt + \beta_2 \cdot Diag$$

Se puede ver con los gráficos cuantil-cuantil, presentado en las Figuras 4.46-4.48, que para estos modelos los  $q_i^{\phi(\lambda)}$  se separan más de la recta  $y = x$ . Es decir, que son menos adecuados para los datos. Además, se observa que los  $q_i^{\phi(\lambda)}$  con  $\lambda = 1$  y  $2/3$  son los que más se ajustan a la recta  $y = x$  pero los más alejados para los otros modelos, sobre todo para el Modelo 1 y Modelo 2. Esta conclusión es la misma que se obtuvo con el estudio de simulación.

Figura 4.46. Gráfico cuantil-cuantil para los datos de la Tabla 4.4 ajustados al Modelo 1

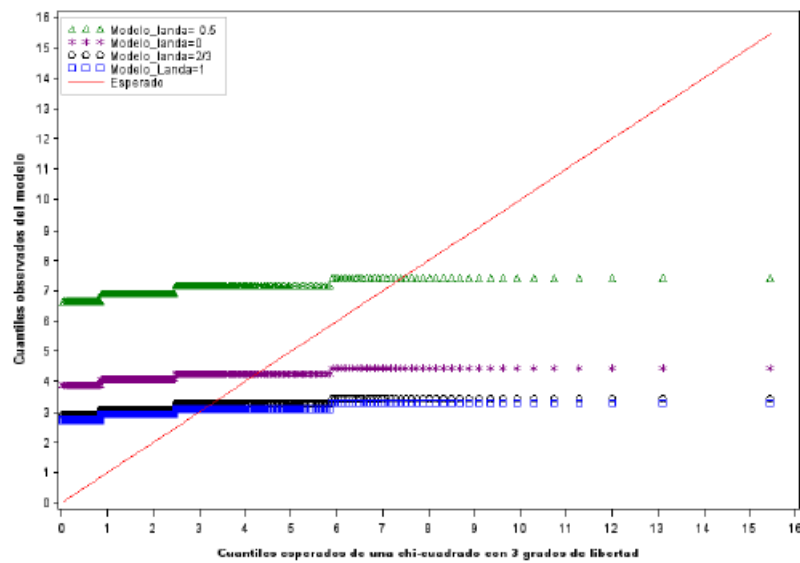


Figura 4.47. Gráfico cuantil-cuantil para los datos de la Tabla 4.4 ajustados al Modelo 2

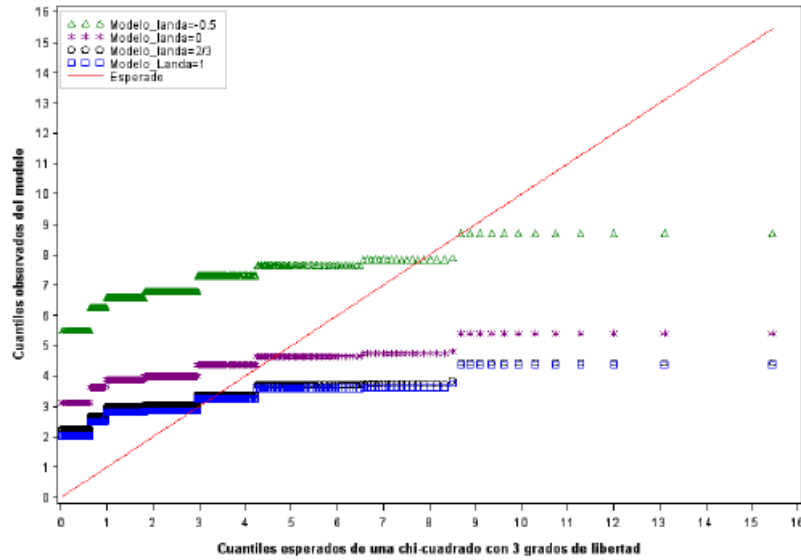
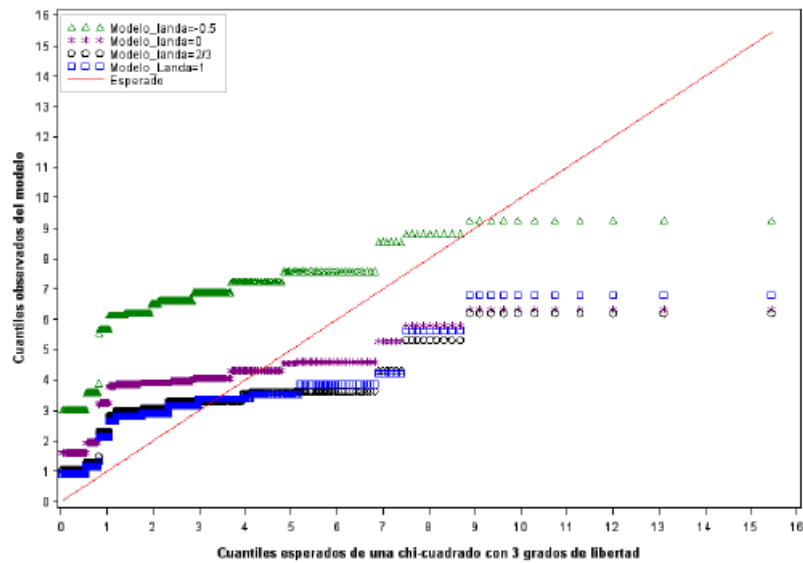


Figura 4.48. Gráfico cuantil-cuantil para los datos de la Tabla 4.4 ajustados al Modelo 3





## Capítulo 5

# Conclusiones y líneas futuras de investigación

La obtención de buenos procedimientos de ajuste para modelos de datos longitudinales redundará en buenos resultados para los estudios en áreas tales como ensayos clínicos, estudios de enfermedades genéticas, estudios epidemiológicos. Análisis de datos correlacionados también son importantes en bioestadística particularmente cuando las medidas se repiten a lo largo del tiempo para el mismo individuo. Por citar algún ejemplo concreto, datos longitudinales se tienen cuando se estudia la eficacia de un medicamento ya que ésto sólo es posible después de un periodo prolongado de administración de dicho medicamento o el estudiar el efecto de la contaminación en un grupo de individuos en los que se miden cada año ciertos síntomas respiratorios a lo largo de un número dado de años, etc.

Entendemos que los procedimientos propuestos en esta tesis son una mejora a algunos de los procedimientos clásicos existentes en la literatura para modelos para datos longitudinales tanto basados en máxima verosimilitud como en ecuaciones de estimación generalizadas. Por otra parte, medidas de diagnóstico para la aproximación EEG no han sido muy estudiadas hasta el momento. En el Capítulo 4 se introducen nuevas herramientas de diagnóstico para detectar “outlying” y puntos influyentes.

En cuanto a líneas futuras de investigación en relación a esta tesis doctoral hay

muchas, destacar las siguientes:

1) Dale [36] propuso una familia de modelos de respuesta bivalente cuyos parámetros se ajustan por máxima verosimilitud. En el Capítulo 2, se propuso una nueva familia de estimadores, los estimadores de mínima  $\phi$ -divergencia, para los Modelos CRG propuestos por Dale [36]. Esta familia de modelos fue generalizada al caso multivariante por Molenberghs y Lesaffre [70]. Una familia amplia de modelos marginales ajustada por máxima verosimilitud es estudiada por Molenberghs y Verbeke [69]. Uno de nuestros más inmediatos objetivos es el de definir un procedimiento de estimación basado en medidas de divergencia para estas generalizaciones del modelo propuesto por Dale [36]. Se tendrían que estudiar sus propiedades asintóticas y de nuevo realizar un amplio estudio de simulación que nos ayudara a encontrar el mejor de los procedimientos de ajuste de estos modelos más generales.

2) La aproximación EEG para analizar datos longitudinales supone que el parámetro de dispersión es común para los diferentes tiempos de observación. Desafortunadamente, lo usual es que esta suposición de un parámetro de dispersión común no se tenga en la práctica. En general, la variable respuesta puede variar a lo largo del tiempo y las condiciones en las que los datos se recogieron pueden diferir en cada tiempo. Park [82] propuso una extensión simple de las EEG que permite diferentes parámetros de dispersión en cada tiempo y fue utilizado para comparar la aproximación EEG con la aproximación de máxima verosimilitud para variables respuesta normales. Park y Shin [102] evalúan el efecto de estos parámetros de dispersión dependientes del tiempo en las estimaciones de los parámetros del modelo. A través de un estudio de simulación estudian la eficiencia de esta extensión de las EEG frente a las EEG propuestas por Liang y Zeger [61]. A esta extensión de las EEG se las puede mejorar haciendo lo mismo que se propuso en el Capítulo 3 como alternativa a las clásicas EEG. Es decir, en lugar de estimar el parámetro de correlación utilizando el residuo de Pearson considerar los residuos  $\phi$ -divergencia. Se haría un estudio de simulación análogo al de Park y Shin [102] con el fin de comprobar si se consigue una mejora a esta extensión de las EEG.

3) Existe bastante literatura en relación al problema de estimación de los parámetros de los modelos lineales generalizados marginales semiparamétricos para datos longitudinales (Liang y Zeger [61]), sin embargo no hay mucho hecho en relación a los

contrastes de hipótesis para estos parámetros. Rao y Scott ([95], [96], [97]) y Roberts, Rao y Kumar [98] estudian ajustes para los estadísticos chi-cuadrado de Pearson y cociente de verosimilitudes usuales en modelos loglineales para tablas de contingencia bajo muestreos complejos. Rotnitzky y Jewell [99] propusieron extensiones a los estadísticos chi-cuadrado de Pearson usuales para llevar a cabo contrastes de hipótesis sobre subconjuntos de parámetros del modelo. Estudian el comportamiento asintótico de los estadísticos de Puntuaciones y Wald generalizados y sus versiones calculadas bajo la suposición que la matriz de correlación de trabajo es verdadera. Sin embargo, una alternativa al procedimiento de estimación de los parámetros consiste en suponer independencia de las observaciones y después ajustar, por ejemplo, los errores estándar estimados para tener en cuenta la dependencia. Para modelos lineales generalizados, este se corresponde con utilizar la estructura de la matriz de trabajo independiente en las EEG. Rotnitzky y Jewell [99] presentan la distribución asintótica del estadístico del cociente de verosimilitudes bajo la suposición de una estructura de independencia en la matriz de trabajo. La idea es proponer nuevos estadísticos basados en divergencias como alternativa al estadístico del cociente de verosimilitudes propuesto por estos autores.

4) Modelos Lineales Mixtos (MLM) son modelos estadísticos que cuantifican las relaciones entre una variable respuesta continua y varias variables explicativas. Un MLM puede incluir tanto parámetros con efecto fijo asociado con una o más variables continuas o categóricas y efectos aleatorios asociados con uno o más factores aleatorios. Esta mezcla de efectos fijos y aleatorios da el nombre a estos modelos. Este modelo es probablemente el más utilizado para analizar datos longitudinales. La utilidad de este modelo para datos longitudinales fue remarcada en un artículo muy citado de Laird y Ware [59]. La literatura acerca de MLM es muy extensa se citará únicamente algunos libros recientes tales como Verbeke y Molenberghs [111], Jiang [54] y West et al. [113]. Sin embargo, hasta ahora el principal interés ha sido cómo ajustar MLM a un conjunto de datos, es decir, en estimar los parámetros desconocidos del modelo. En la investigación que se pretende desarrollar nos centraremos en alguno de los problemas de contraste para MLM que no han sido muy estudiados hasta el momento. Contrastes de hipótesis en los que los parámetros covarianza tienen valores que caen en la frontera del espacio paramétrico surgen con frecuencia cuando se quiere contrastar si un determinado efecto aleatorio de un MLM se

debería mantener en el modelo o no. En lugar de contrastar directamente sobre los efectos aleatorios se contrasta si las correspondientes varianzas y covarianzas son iguales a cero. En el caso en el que tengamos un único efecto aleatorio en un modelo, se puede estar interesado en contrastar la hipótesis nula de que el efecto aleatorio se puede eliminar. Self y Liang [101], Stram y Lee [106], y Verbeke y Molenberghs [111] demostraron que el estadístico del cociente de verosimilitudes, en este caso tiene una distribución asintótica que es una mixtura de dos chi-cuadrado de 0 y 1 grados de libertad, respectivamente. En el caso en el que tengamos dos efectos aleatorios en un modelo y se quiera contrastar que uno de ellos se puede eliminar, es necesario contrastar si la varianza para el efecto aleatorio dado que se desea contrastar y la covarianza asociada de los dos efectos aleatorios son ambos iguales a cero. La distribución asintótica del estadístico del cociente de verosimilitudes en este caso es una mixtura de dos chi-cuadrado con grados de libertad 1 y 2, respectivamente, (Verbeke y Molenberghs [111]). En los trabajos citados anteriormente suponen que el vector de la variable respuesta se puede particionar en subvectores independientes e idénticamente distribuidos y el número de subvectores independientes tienden a infinito. Esta restricción es eliminada en el trabajo de Crainiceanu y Ruppert [31]. Estos autores consideran el problema de contrastar hipótesis nulas que incluyen restricciones en la componente varianza. El contraste que proponen no es estándar ya que los parámetros bajo la hipótesis nula están en la frontera del espacio paramétrico y también porque las variables respuesta no son independientes bajo la alternativa. Molina y Morales [72] generalizaron el estadístico del cociente de verosimilitudes para contrastar hipótesis generales en modelos de regresión lineal mixta utilizando la divergencia de Rényi. En el estudio de simulación que llevan a cabo concluyen que algunos de los miembros de la nueva familia de estadísticos son mejores que el clásico test del cociente de verosimilitudes. Una nueva línea de investigación sería estudiar estadísticos basados en la divergencia de Rényi en el contexto más complejo propuesto por Crainiceanu y Ruppert [31] para contrastar hipótesis que incluyen restricciones sobre la componente varianza en MLM y obtener tanto la distribución asintótica como la muestral finita de los nuevos estadísticos sin restringir y restringidos. Todos estos resultados teóricos se complementarían con un estudio de simulación con el objetivo de estudiar el comportamiento de los nuevos estadísticos bajo diferentes alternativas. Finalmente, modelos aditivos de spline-

penalizados descritos en Marx y Eilers [66], Ruppert y Carroll [100] y Aerts et al. [1] tienen la ventaja de sólo necesitar un pequeño conjunto de funciones basadas en splines para cada variable explicativa y pueden ser representados como modelos mixtos (Brumback et al. [22]). Contrastes para simplificar suposiciones, tales como el no efecto de una variable explicativa o un efecto lineal, se puede reducir a contrastar que las componentes de la varianza son cero. Crainiceanu et al. [32] proponen estadísticos del cociente de verosimilitudes para contrastar regresión polinomial frente a alternativas generales modelizadas mediante splines penalizados. Otro de los objetivos sería abordar también este tema basándonos en los resultados teóricos que hayamos obtenidos previamente.





# Anexo

## Primeras derivadas parciales de la Divergencia de Csiszár para Modelos CRG

Sea  $Y_{ijk}$ ,  $i = 1, \dots, r$ ;  $j = 1, \dots, c$ , una multinomial de parámetros  $n_k = \sum_{i=1}^r \sum_{j=1}^c Y_{ijk}$  observaciones y probabilidades  $\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})$  con

$$\boldsymbol{\theta} = (\alpha_1, \beta_1, \alpha_2, \beta_2, \boldsymbol{\Delta}, \alpha_a, \beta_a, \boldsymbol{\delta}, \boldsymbol{\gamma}).$$

En lo que sigue calculamos las primeras derivadas parciales de

$$d_k^\phi = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^c n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta}) \phi \left( \frac{y_{ijk}}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \right)$$

con respecto a  $\boldsymbol{\theta}$ .

Sea

$$\zeta(i, j) = \frac{1}{n} \left[ n_k \phi \left( \frac{y_{ijk}}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \right) - \frac{y_{ijk}}{\pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \phi' \left( \frac{y_{ijk}}{n_k \pi_{ij}(\mathbf{x}_k; \boldsymbol{\theta})} \right) \right]$$

entonces

$$\frac{\partial d_k^\phi}{\partial F_{ijk}} = \zeta(i, j) - \zeta(i+1, j) - \zeta(i, j+1) + \zeta(i+1, j+1)$$

$$\frac{\partial d_k^\phi}{\partial F_{rjk}} = \zeta(r, j) - \zeta(r, j+1)$$

$$\frac{\partial d_k^\phi}{\partial F_{ick}} = \zeta(i, c) - \zeta(i+1, c)$$

$$\frac{\partial d_k^\phi}{\partial F_{rck}} = \zeta(r, c)$$

con  $i = 1, \dots, r$ ;  $j = 1, \dots, c$ .

Puesto que

$$\frac{\partial d_k^\phi}{\partial \eta_{ik}} = \frac{\partial d_k^\phi}{\partial F_{ick}} + \sum_{j=1}^{c-1} \frac{\partial d_k^\phi}{\partial F_{ijk}} \frac{\partial F_{ijk}}{\partial \eta_{ik}}, \quad i = 1, \dots, r-1$$

a continuación se calculan las derivadas de  $F_{ijk}$  con respecto de  $\eta_{ik}$ .

Al ser

$$\begin{aligned} F_{ijk} &= \frac{1}{2} (\psi_{ijk} - 1)^{-1} [1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1) - S(\eta_{ik}, \xi_{jk}, \psi_{ijk})] \\ &= \frac{1}{2} (\psi_{ijk} - 1)^{-1} [1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1) - \\ &\quad \sqrt{[1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1)]^2 + 4\psi_{ijk} (1 - \psi_{ijk}) \eta_{ik} \xi_{jk}}] \end{aligned}$$

se tiene que

$$\begin{aligned} \frac{\partial F_{ijk}}{\partial \eta_{ik}} &= \frac{1}{2} (\psi_{ijk} - 1)^{-1} [(\psi_{ijk} - 1) - \\ &\quad \frac{2[1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1)] (\psi_{ijk} - 1) + 4\psi_{ijk} (1 - \psi_{ijk}) \xi_{jk}}{2\sqrt{[1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1)]^2 + 4\psi_{ijk} (1 - \psi_{ijk}) \eta_{ik} \xi_{jk}}}] = \\ &= \frac{1}{2} \left[ 1 - \frac{1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1) - 2\psi_{ijk} \xi_{jk}}{\sqrt{[1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1)]^2 + 4\psi_{ijk} (1 - \psi_{ijk}) \eta_{ik} \xi_{jk}}} \right] \\ &= \frac{1}{2} [1 - S_{ijk}^{-1} (1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1) - 2\psi_{ijk} \xi_{jk})] \\ &= \frac{1}{2} [1 - S_{ijk}^{-1} (1 + (\eta_{ik} - \xi_{jk}) \psi_{ijk} - \eta_{ik} - \xi_{jk})]. \end{aligned}$$

Poniendo  $\Delta_{ij} = \log(\psi_{ijk})$ , se tiene

$$\frac{\partial F_{ijk}}{\partial \eta_{ik}} = \frac{1}{2} \left\{ 1 - S_{ijk}^{-1} [1 + \exp(\Delta_{ij}) (\eta_{ik} - \xi_{jk}) - \eta_{ik} - \xi_{jk}] \right\}.$$

De forma análoga,

$$\frac{\partial d_k^\phi}{\partial \xi_{jk}} = \frac{\partial d_k^\phi}{\partial F_{rjk}} + \sum_{i=1}^{r-1} \frac{\partial d_k^\phi}{\partial F_{ijk}} \frac{\partial F_{ijk}}{\partial \xi_{jk}}, \quad j = 1, \dots, c-1$$

donde

$$\begin{aligned}
\frac{\partial F_{ijk}}{\partial \xi_{jk}} &= \frac{1}{2} (\psi_{ijk} - 1)^{-1} [(\psi_{ijk} - 1) \\
&\quad - \frac{2 [1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1)] (\psi_{ijk} - 1) + 4\psi_{ijk} (1 - \psi_{ijk}) \eta_{ik}}{2\sqrt{[1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1)]^2 + 4\psi_{ijk} (1 - \psi_{ijk}) \eta_{ik} \xi_{jk}}}] \\
&= \frac{1}{2} \left[ 1 - \frac{1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1) - 2\psi_{ijk} \eta_{ik}}{\sqrt{[1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1)]^2 + 4\psi_{ijk} (1 - \psi_{ijk}) \eta_{ik} \xi_{jk}}} \right] \\
&= \frac{1}{2} \left[ 1 - S_{ijk}^{-1} (1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1) - 2\psi_{ijk} \eta_{ik}) \right].
\end{aligned}$$

Teniendo en cuenta, como el caso anterior que  $\psi_{ijk} = \exp \Delta_{ij}$ , podemos escribir la derivada anterior como:

$$\begin{aligned}
\frac{\partial F_{ijk}}{\partial \xi_{jk}} &= \frac{1}{2} \left[ 1 - S_{ijk}^{-1} (1 + (\eta_{ik} + \xi_{jk}) (\exp(\Delta_{ij}) - 1) - 2 \exp(\Delta_{ij}) \eta_{ik}) \right] \\
&= \frac{1}{2} \left[ 1 - S_{ijk}^{-1} (1 + \exp(\Delta_{ij}) (\xi_{jk} - \eta_{ik}) - \eta_{ik} - \xi_{jk}) \right].
\end{aligned}$$

Por otra parte, también es necesario calcular la derivada de  $F_{ijk}$  con respecto a  $\Delta_{ij}$ .

Utilizando la regla de la cadena se tiene:

$$\frac{\partial F_{ijk}}{\partial \Delta_{ij}} = \frac{\partial F_{ijk}}{\partial \psi_{ijk}} \frac{\partial \psi_{ijk}}{\partial \Delta_{ij}}$$

donde

$$\begin{aligned}
\frac{\partial F_{ijk}}{\partial \psi_{ijk}} &= \frac{1}{2} \left( \frac{-1}{(\psi_{ijk} - 1)^2} \right) [1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1) \\
&\quad - \sqrt{[1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1)]^2 + 4\psi_{ijk} (1 - \psi_{ijk}) \eta_{ik} \xi_{jk}}] + \\
&\quad + \frac{1}{2} (\psi_{ijk} - 1)^{-1} [(\eta_{ik} + \xi_{jk}) - \\
&\quad \frac{2 [1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1)] (\eta_{ik} + \xi_{jk}) + 4 (1 - \psi_{ijk}) \eta_{ik} \xi_{jk} - 4\psi_{ijk} \eta_{ik} \xi_{jk}}{2\sqrt{[1 + (\eta_{ik} + \xi_{jk}) (\psi_{ijk} - 1)]^2 + 4\psi_{ijk} (1 - \psi_{ijk}) \eta_{ik} \xi_{jk}}}] =
\end{aligned}$$

$$\begin{aligned}
&= \frac{-1}{2(\psi_{ijk} - 1)^2} [1 + (\eta_{ik} + \xi_{jk})(\psi_{ijk} - 1) - S_{ijk}] + \\
&+ \frac{1}{2}(\psi_{ijk} - 1)^{-1} [(\eta_{ik} + \xi_{jk}) \\
&- S_{ijk}^{-1} ([1 + (\eta_{ik} + \xi_{jk})(\psi_{ijk} - 1)] (\eta_{ik} + \xi_{jk}) + 2(1 - \psi_{ijk}) \eta_{ik} \xi_{jk} - 2\psi_{ijk} \eta_{ik} \xi_{jk})] \\
&= \frac{-1}{2(\psi_{ijk} - 1)^2} - \frac{(\eta_{ik} + \xi_{jk})}{2(\psi_{ijk} - 1)} + \frac{S_{ijk}}{2(\psi_{ijk} - 1)^2} + \frac{(\eta_{ik} + \xi_{jk})}{2(\psi_{ijk} - 1)} - \\
&- \frac{1}{2S_{ijk}(\psi_{ijk} - 1)} \left\{ (\eta_{ik} + \xi_{jk}) + (\eta_{ik} + \xi_{jk})^2 (\psi_{ijk} - 1) + 2(1 - 2\psi_{ijk}) \eta_{ik} \xi_{jk} \right\} = \\
&= \frac{-S_{ijk} + S_{ijk}^2 - (\eta_{ik} + \xi_{jk})(\psi_{ijk} - 1) - (\eta_{ik} + \xi_{jk})^2 (\psi_{ijk} - 1)^2}{2(\psi_{ijk} - 1)^2 S_{ijk}} \\
&+ \frac{4\psi_{ijk}(\psi_{ijk} - 1) \eta_{ik} \xi_{jk} - 2\eta_{ik} \xi_{jk} (\psi_{ijk} - 1)}{2(\psi_{ijk} - 1)^2 S_{ijk}} \\
&= \frac{-S_{ijk} + 1 + (\eta_{ik} + \xi_{jk})(\psi_{ijk} - 1)^2 + 4\psi_{ijk}(1 - \psi_{ijk}) \eta_{ik} \xi_{jk} - (\eta_{ik} + \xi_{jk})(\psi_{ijk} - 1)}{2(\psi_{ijk} - 1)^2 S_{ijk}} \\
&+ \frac{-(\eta_{ik} + \xi_{jk})^2 (\psi_{ijk} - 1)^2 + 4\psi_{ijk}(\psi_{ijk} - 1) \eta_{ik} \xi_{jk} - 2\eta_{ik} \xi_{jk} (\psi_{ijk} - 1)}{2(\psi_{ijk} - 1)^2 S_{ijk}} = \\
&= \frac{-S_{ijk} + 1 - (\eta_{ik} + \xi_{jk})(\psi_{ijk} - 1) - 2\eta_{ik} \xi_{jk} (\psi_{ijk} - 1)}{2(\psi_{ijk} - 1)^2 S_{ijk}} = \\
&= \left( 2(\psi_{ijk} - 1)^2 S_{ijk} \right)^{-1} (1 - S_{ijk} + (\psi_{ijk} - 1)(\eta_{ik} + \xi_{jk} - 2\eta_{ik} \xi_{jk}))
\end{aligned}$$

y

$$\frac{\partial \psi_{ijk}}{\partial \Delta} = \exp \Delta_{ij}$$

por tanto

$$\frac{\partial F_{ijk}}{\partial \Delta_{ij}} = \exp(\Delta_{ij}) \left( 2(\exp(\Delta_{ij}) - 1)^2 S_{ijk} \right)^{-1} (1 - S_{ijk} + (\exp(\Delta_{ij}) - 1)(\eta_{ik} + \xi_{jk} - 2\eta_{ik} \xi_{jk})).$$

A que continuación se calcula la derivada de la función  $d_k^\phi$  con respecto de cada una de las componentes de  $\theta = (\alpha_1, \beta_1, \alpha_2, \beta_2, \Delta, \alpha_a, \beta_a, \delta, \gamma)$ .

En primer lugar calculamos la derivada de  $\frac{\partial d_k^\phi}{\partial \alpha_{1i}}$ , teniendo en cuenta la regla de la cadena podemos expresar la derivada como:

$$\frac{\partial d_k^\phi}{\partial \alpha_{1i}} = \frac{\partial d_k^\phi}{\partial \eta_{ik}} \frac{\partial \eta_{ik}}{\partial \alpha_{1i}}$$

Empezamos por calcular  $\frac{\partial \eta_{ik}}{\partial \alpha_{1i}}$ . Al ser  $\eta_{ik} = \frac{\exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1)}{1 + \exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1)}$ ,  $k = 1, \dots, r-1$

entonces

$$\begin{aligned} \frac{\partial \eta_{ik}}{\partial \alpha_{1i}} &= \frac{\exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1) (1 + \exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1)) - \exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1) \exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1)}{(1 + \exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1))^2} = \\ &= \frac{\exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1)}{(1 + \exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1))^2} = \frac{\exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1)}{1 + \exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1)} \frac{1}{1 + \exp(\alpha_{1k} - \mathbf{x}_k^T \boldsymbol{\theta}_1)} = \\ &= \eta_{ik} (1 - \eta_{ik}) \end{aligned}$$

por tanto

$$\frac{\partial d_k^\phi}{\partial \alpha_{1i}} = \frac{\partial d_k^\phi}{\partial \eta_{ik}} \eta_{ik} (1 - \eta_{ik}), \quad i = 1, \dots, r-1.$$

En segundo lugar calculamos  $\frac{\partial d_k^\phi}{\partial \boldsymbol{\theta}_1}$ , teniendo en cuenta la regla de la cadena

$$\frac{\partial d_k^\phi}{\partial \boldsymbol{\theta}_1} = \frac{\partial d_k^\phi}{\partial \eta_{ik}} \frac{\partial \eta_{ik}}{\partial \boldsymbol{\theta}_1}.$$

Al ser

$$\begin{aligned} \frac{\partial \eta_{ik}}{\partial \boldsymbol{\theta}_1} &= \frac{-\exp(\alpha_{1i} - \mathbf{x}_k^T \boldsymbol{\theta}_1) \mathbf{x}_k^T (1 + \exp(\alpha_{1i} - \mathbf{x}_k^T \boldsymbol{\theta}_1)) + \exp(\alpha_{1i} - \mathbf{x}_k^T \boldsymbol{\theta}_1) \mathbf{x}_k^T \exp(\alpha_{1i} - \mathbf{x}_k^T \boldsymbol{\theta}_1)}{(1 + \exp(\alpha_{1i} - \mathbf{x}_k^T \boldsymbol{\theta}_1))^2} = \\ &= \frac{-\mathbf{x}_k^T \exp(\alpha_{1i} - \mathbf{x}_k^T \boldsymbol{\theta}_1)}{(1 + \exp(\alpha_{1i} - \mathbf{x}_k^T \boldsymbol{\theta}_1))^2} = -\mathbf{x}_k^T \eta_{ik} (1 - \eta_{ik}) \end{aligned}$$

entonces

$$\frac{\partial d_k^\phi}{\partial \theta_1} = \sum_{i=1}^{r-1} \frac{\partial d_k^\phi}{\partial \eta_{ik}} (-\mathbf{x}_k^T \eta_{ik} (1 - \eta_{ik})) = -\mathbf{x}_k^T \sum_{i=1}^{r-1} \frac{\partial d_k^\phi}{\partial \eta_{ik}} \frac{\partial \eta_{ik}}{\partial \alpha_{1i}} = -\mathbf{x}_k^T \sum_{i=1}^{r-1} \frac{\partial d_k^\phi}{\partial \alpha_{1i}}.$$

En tercer lugar, calculamos  $\frac{\partial d_k^\phi}{\partial \alpha_{2j}}$  teniendo en cuenta una vez más la regla de la cadena

$$\frac{\partial d_k^\phi}{\partial \alpha_{2j}} = \frac{\partial d_k^\phi}{\partial \xi_{jk}} \frac{\partial \xi_{jk}}{\partial \alpha_{2j}}.$$

Empezamos por calcular  $\frac{\partial \xi_{jk}}{\partial \alpha_{2j}}$ . Al ser  $\xi_{jk} = \frac{\exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2)}{1 + \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2)}$  con  $j = 1, \dots, c-1$

se tiene que

$$\begin{aligned} \frac{\partial \xi_{jk}}{\partial \alpha_{2j}} &= \frac{\exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2) (1 + \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2)) - \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2) \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2)}{(1 + \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2))^2} \\ &= \frac{\exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2)}{(1 + \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2))^2} = \frac{\exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2)}{1 + \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2)} \frac{1}{1 + \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2)} \\ &= \xi_{jk} (1 - \xi_{jk}); \quad j = 1, \dots, c-1 \end{aligned}$$

entonces

$$\frac{\partial d_k^\phi}{\partial \alpha_{2j}} = \frac{\partial d_k^\phi}{\partial \xi_{jk}} \frac{\partial \xi_{jk}}{\partial \alpha_{2j}} = \xi_{jk} (1 - \xi_{jk}) \frac{\partial d_k^\phi}{\partial \xi_{jk}}; \quad j = 1, \dots, c-1.$$

En cuarto lugar, calculamos  $\frac{\partial d_k^\phi}{\partial \boldsymbol{\theta}_2}$  siguiendo el mismo procedimiento que en los casos

precedentes, empezamos por calcular:

$$\begin{aligned} \frac{\partial \xi_{jk}}{\partial \boldsymbol{\theta}_2} &= \frac{-\mathbf{x}_k^T \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2) (1 + \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2)) + \mathbf{x}_k^T \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2) \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2)}{(1 + \exp(\alpha_{2j} - \mathbf{x}_k^T \boldsymbol{\theta}_2))^2} \\ &= -\mathbf{x}_k^T \frac{\partial \xi_{jk}}{\partial \alpha_{2j}} \\ \frac{\partial d_k^\phi}{\partial \boldsymbol{\theta}_2} &= \sum_{j=1}^{c-1} \frac{\partial d_k^\phi}{\partial \xi_{jk}} \frac{\partial \xi_{jk}}{\partial \boldsymbol{\theta}_2} = -\mathbf{x}_k^T \sum_{j=1}^{c-1} \frac{\partial d_k^\phi}{\partial \xi_{jk}} \frac{\partial \xi_{jk}}{\partial \alpha_{2j}} = -\mathbf{x}_k^T \sum_{j=1}^{c-1} \frac{\partial d_k^\phi}{\partial \alpha_{2j}}. \end{aligned}$$

Además

$$\frac{\partial d_k^\phi}{\partial \Delta_{ij}} = \frac{\partial d_k^\phi}{\partial F_{ijk}} \frac{\partial F_{ijk}}{\partial \Delta_{ij}}.$$

A continuación se especifican las derivadas de  $d_k^\phi$  con respecto a  $\Delta$ ,  $\alpha_{ia}$ ,  $\beta_{ja}$ ,  $\delta_{ij}$  y  $\gamma$ , teniendo en cuenta que  $\psi_{ijk} = \exp \Delta_{ij}$  y que  $\log(\psi_{ijk}) = \Delta + \alpha_{ia} + \beta_{ja} + \delta_{ij} - \gamma^T \mathbf{x}$

$$\frac{\partial d_k^\phi}{\partial \Delta} = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} \frac{\partial d_k^\phi}{\partial \Delta_{ij}}$$

$$\frac{\partial d_k^\phi}{\partial \alpha_{ia}} = \sum_{j=1}^{c-1} \frac{\partial d_k^\phi}{\partial \Delta_{ij}} \frac{\partial \Delta_{ij}}{\partial \alpha_{ia}} = \sum_{j=1}^{c-1} \frac{\partial d_k^\phi}{\partial \Delta_{ij}}, \quad i = 1, \dots, r-2$$

$$\frac{\partial d_k^\phi}{\partial \beta_{ja}} = \sum_{i=1}^{r-1} \frac{\partial d_k^\phi}{\partial \Delta_{ij}} \frac{\partial \Delta_{ij}}{\partial \beta_{ja}} = \sum_{i=1}^{r-1} \frac{\partial d_k^\phi}{\partial \Delta_{ij}}, \quad j = 1, \dots, c-2$$

$$\frac{\partial d_k^\phi}{\partial \delta_{ij}} = \frac{\partial d_k^\phi}{\partial \Delta_{ij}} \frac{\partial \Delta_{ij}}{\partial \delta_{ij}} = \frac{\partial d_k^\phi}{\partial \Delta_{ij}}$$

$$\frac{\partial d_k^\phi}{\partial \gamma} = -\mathbf{x}_k \frac{\partial d_k^\phi}{\partial \Delta}$$





# Bibliografía

- [1] Aerts, M., Claeskens, G. y Wand, M.P. (2002). Some theory for penalized spline additive models. *J. Statist. Plann. Inference*, 103, 455-470.
- [2] Agresti, A. (2002). *Categorical Data Analysis*. 2nd edition. Wiley.
- [3] Agresti, A. y Lang, J.B. (1994). A Proportional Odds Model with Subject-Specific Effects for Repeated Ordered Categorical Responses. *Biometrika*, 80, (3), 527-534
- [4] Airy, G.B. (1861). *On the algebraical and numerical theory of errors of observations and the combination of observations*. Cambridge: MacMillan and co.
- [5] Albert, P. y McShane, L. (1995). A generalized estimating equations approach for spatially data via the EM algorithm (with discussion). *Journal of the Royal Society Series B*, 39, 1-38.
- [6] Altham, P. (1978). Two generalizations of the Binomial distribution. *App. Statist.* 27, 162-167.
- [7] Anderson, T.W. y Goodman, L.A. (1957). Statistical inference about Markov chains. *Annals of mathematical statistics*, 28, 89-110.
- [8] Anscombe, F. J. (1953). Contribution to the discussion of H. Hotelling's paper. *J. Roy. Statist. Soc. B15*, 229-230.
- [9] Anscombe, F.J. (1981). *Computing in Statistical Science through APL*. New York: Springer-Verlag.

- 
- [10] Atkinson, A.C. (1985). *Plots, Transformations and Regressions*. Oxford: Oxford Statistical Science Series.
- [11] Azzalini, A. (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika*, 81, 767-775.
- [12] Bahadur, R.R. (1961). *A representation of the joint distribution of responses to  $n$  dichotomous items*. In: Solomon H, ed. Studies in item analysis and prediction. Stanford Mathematical Studies in the Social Sciences VI, Stanford, CA: Stanford University Press.
- [13] Baltes, P. B. y Nesselroade, J. R. (1972). Cultural change and adolescent personality development: An application of longitudinal sequences. *Developmental Psychology*, 7, 244- 256.
- [14] Ballinger, G. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods*, 7, (2), 127-150.
- [15] Barnhart, H.X. y Williamson, J.M. (1998). Goodness-of-fit test for GEE modeling with binary responses. *Biometrics* 54, 720-729.
- [16] Basu, A. y Basu, S. (1998). Penalized minimum disparity methods for multinomial models. *Statistica Sinica*, 8, 841-860.
- [17] Berkson, J. (1955). Maximum likelihood and minimum  $\chi^2$  estimates of the logistic function. *Journal of American Statistical Association*, 50, 130-162.
- [18] Besley, D.A., Kuh, E., y Welsch, R.E. (1980). *Regression Diagnostics*. New York: Wiley.
- [19] Billingsley, P. (1961). *Statistical inference for Markov processes*. Chicago : University of Chicago Press, 75.
- [20] Breslow, N.E. y Clayton, D.G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88, (421), 9-25.

- 
- [21] Breslow, N.E. y Day, N.E. (1987). *Statistical Methods in Cancer Research*. Vol. II, The Design and Analysis of Cohort Studies (IARC Scientific Publication No. 82). Lyon, France: International Agency for Research on Cancer.
- [22] Brumback, B.A., Ruppert, D. y Wand, M.P. (1999). Variable Selection and Function Estimation in Additive Nonparametric Regression Using a Data-Based. *Journal of the American Statistical Association*, 94, (447), 194-797.
- [23] Carey, V., Zeger, S.L. y Diggle, P. (1993). Modelling Multivariate Binary Data with Alternating Logistic Regressions. *Biometrika*, 80, (3), 517-526.
- [24] Cessie, S. y Van Houwelingen, J.C. (1984). Logistic Regression for Correlated Binary Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43, (1), 95-108.
- [25] Chang, Y.C. (2000). Residuals analysis of the generalized linear models for longitudinal data. *Statistics in Medicine*, 19, 1277-1293.
- [26] Chatterjee, S., y Hadi, A.S. (1986). Influential Observations, High Leverage, and Outliers in Linear Regression. *Statistical Science*, 1, (3), 379-416.
- [27] Cook, R.D. y Weisberg, S. (1982). *Residuals and Influence in Regression*, New York: Chapman & Hall.
- [28] Cox, D.R. (1958). The regression analysis of binary sequences. *J.R. Statist. Soc. B*, 20, 215-242.
- [29] Cox, D.R. (1970). *The Analysis of Binary Data*. London: Methuen.
- [30] Cox, D.R. y Hinkley, D.V. (1974). *Theoretical Statistics*. London : Chapman and Hall.
- [31] Crainiceanu, C.M. y Ruppert, D. (2004). Likelihood Ratio Test in Linear Mixed Models with One Variance Component. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66, 1, 165-185.
- [32] Crainiceanu, C., Ruppert, D., Claeskens, G. y Wand, M.P. (2005). Exact likelihood ratio test for penalised splines. *Biometrika*, 92, (1), 91-103.

- 
- [33] Cressie, N. y Read, T.R.C. (1984). Multinomial goodness-of-fit test. *Journal of the Royal Statistical Society, Serie B*, 46, 440-464.
- [34] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2, 51-66.
- [35] Dahmen, G. y Ziegler, A. (2004). Generalized estimating equations in controlled clinical trials: hypotheses testing. *Biom*; 46, (2), 214-32.
- [36] Dale, J.R. (1986). Global Cross-Ratio Models for Bivariate, Discrete, Ordered Responses. *Biometrics*, 42, (4), 909-917.
- [37] Diggle, P. J. (1988). An approach to the analysis of repeated measures. *Biometrics* 44, 959-971.
- [38] Diggle, P.J., Liang K.Y. y Zeger, S.L. (1994). *Analysis of Longitudinal Data*. Oxford: Oxford Statistical Science Publications.
- [39] Diggle, P. J, Liang, K.-Y. y Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford Statistical Science Series 25. Oxford University Press.
- [40] Fahrmeir, L. y Tutz, G. (1994), (2001): *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer Series in Statistics. Springer Verlag. New-York Berlin Heidelberg.
- [41] Fitzmaurice, G.M. y Laird, N.M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, 80, 141-151.
- [42] Fitzmaurice, G.M., Laird, N.M. y Rotnitzky, A.G. (1993). Regression Models for Discrete Longitudinal Responses: Rejoinder. Source: *Statist. Sci.* 8, (3), 306-309.
- [43] Fitzmaurice, G., Laird, N., Tosteson, T. (1996). *Polynomial Exponential Models for Clustered Binary Outcomes*. Unpublished manuscript.
- [44] Flemming, W. (1977). *Functions of Several Variables (Second Edition)*, Springer-Verlag, New York.

- 
- [45] Glonek, G.F.V. y McCullagh, P. (1995). Multivariate Logistic Models. *Journal of the Royal Statistical Society. Serie B (Methodological)*, 57, I(3), 533-546.
- [46] Goldstein, H. (1979). *The design and analysis of longitudinal studies: their role in the measurement of change*. London. Academic Press.
- [47] Goodman, L.A. (1974). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association* 74, 537-552.
- [48] Hardin, J.W. y J. M. Hilbe. (2003). *Generalized Estimating Equations*. Boca Raton, FL: Chapman & Hall/CRC.
- [49] Hawkins, D.M. y Wixley, R.A.J. (1986). A note of the transformation of Chi-squared variables to normality. *Amer. Statist.* 40, 296-298.
- [50] Heagerty, P.J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics* 55, 688-698.
- [51] Heagerty, P.J. y Zeger, S.L. (2000). Marginalized multilevel models and likelihood inference. *Source: Statist. Sci.* 15, (1), 1-26.
- [52] Hedeker, D. y Gibbons, R. D. (2006). *Longitudinal Data Analysis*. Wiley Series in Probability and Statistics. Wiley.
- [53] Horton, N. J. y Lipstiz, S. R. (1999). Review of software to fit generalized estimating equation regression models. *The American Statistician*.
- [54] Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer Serier in Statistics. Springer.
- [55] Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., Lehnen, R.G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics* 33, 133-158.
- [56] Koch, G.G., Carr, G.J., Amara, I.A., Stokes, M.E. y Uryniak, T.J. (1990). *Categorical data analysis, in Statistical Methodology in the Pharmaceutical Sciences*. Ed. D.A. Berry, New York: Marcel Dekker Inc., 391-475.

- 
- [57] Kullback, S. y Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, (1), 79-86.
- [58] Kupper, L.L. y Haseman, J.K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics*, 34, 69-76.
- [59] Laird, N. y Ware, J. (1982). Random-Effects Models for Longitudinal Data. *Biometrics* 38, 963-974.
- [60] Leppik, I.E., Dreifuss, E., Bowman-Cloyd, T. Santilli. (1985). A double-blind crossover evaluation of progabide in partial seizures. *Neurology*, 35, 285.
- [61] Liang, K.Y. y Zeger, S. L. (1986). Longitudinal data analisis using generalized linear models. *Biometrika*, 73, 13-22.
- [62] Liang, K.Y., Zeger, S. L. y Qaqish, B. (1992). Multivariate regression analyses for categorical data. *J. Roy. Statist. Soc. Ser. B* 54, 3-40.
- [63] Littell, R.C., Milliken, G.A., Stroup, W.W. y Wolfinger, R.D. (1996). *SAS System for Mixed Models*. SAS Institute Inc., Cary, NC, USA.
- [64] Lipstz, S.R., Laird, N.M. y Harrington, D.P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika* 78, 153-160.
- [65] Mardia, K.V. (1970). *Families of Bivariate Distributions*. London: Griffin.
- [66] Marx, B.D. y Eilers, P.H.C. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics and Data Analysis*, 28, 193-209.
- [67] McCullagh, P. (1980). Regression models for ordinal data (with Discussion). *Journal of the Royal Statistical Society, Series B* 42, 109-142.
- [68] McCullagh, P. y Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman y Hall, London.
- [69] Molenberghs, G. y Verdeke, G. (2005). *Models for discrete longitudinal data*. Springer, New York. 687.

- 
- [70] Molenberghs, G. y Lesaffre, E. (1994). Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *J. Amer. Statist. Assoc.*, (89), 633-644.
- [71] Molenberghs G. y Ryan L.M. (1999). Likelihood inference for clustered multivariate binary data. *Environmetrics* 10, 279-300 .
- [72] Molina, I. y Morales, D. (2007). Rényi statistics for testing hypotheses in mixed linear regression models. *Journal of Statistical Planning and Inference*, 137, 87-102.
- [73] Nelder, J. A. y Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*. 135, (3), 370-384.
- [74] Nesselroade, J. R., y Baltes, P. B. (Eds.). (1979). *Longitudinal research in the study of behavior and development*. New York: Academic Press.
- [75] Neuhaus, J.M. , Kalbfleisch, J.M. y Hauck, W.W. (1991). A Comparison of Cluster-Specific and Population-Averaged Approaches for Analyzing Correlated Binary Data. *Source: International Statistical Review / Revue Internationale de Statistique*, 59, (1), 25-35.
- [76] Oh, S., Carriere, K.C. y Park, T. (2008). Model diagnostic plots for repeated measures data using the generalized estimating equations approach. *Computational Statistics and Data Analysis*, 53, 222-232.
- [77] Palmgren, J. (1989). Regression models for bivariate binary responses. *Technical Report 101. Department of Biostatistics*, University of Washington, Seattle.
- [78] Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics*, **57**, 120-125.
- [79] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*. Ed. Chapman & Hall.
- [80] Pardo, L. y Pardo, M.C. (2003). Minimum power-divergence estimator in three-way contingency tables. *Journal of Statistical Computation and simulation*. 73, (11), 819-831.



- 
- [81] Pardo, J.A., Pardo, L. y Pardo, M.C. (2006). Testing in logistic regression models based on  $\phi$ -divergence measure. *Journal of Statistical Planning and Inference*, 136, 982-1006.
- [82] Park, T. (1993). A comparison of the generalized estimating equation approach with the maximum likelihood approach for repeated measurements. *Statistics in Medicine*. 12, (18), 1723-1732.
- [83] Park, C.G., Park, T. y Shin, D.W. (1996). A simple method for generating correlated binary variates. *The American Statistician*, 50, (4), 306-310.
- [84] Park, T., Davis, C. S. y Li, N. (1998). Alternative GEE estimation procedures for discrete longitudinal data. *Computational Statistics & Data Analysis* 28, 243-256.
- [85] Park, C. G. y Shin, D.W. (1998). An algorithm for generating correlated random variables in a class of infinitely divisible distributions. *Journal Statist. Comput. Simul.*, 61, 127-139.
- [86] Park, T. y Lee, S.Y. (2004). Model diagnostic plots for repeated measures data. *Biom. J.* 46, 441-452.
- [87] Paik, M.C. (1992). Parametric variance function estimation for nonnormal repeated measurement data. *Biometrics* 48, 19-30.
- [88] Pearson, K. y Heron, D. (1913). On theories of association. *Biometrika* 9, 159-315.
- [89] Plackett, R.L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association* 60, 516-522.
- [90] Pregibon, D. (1981). Logistic Regression Diagnostics. *The Annals of Statistics*. 9, (4), 705-724.
- [91] Preisser, J.S. y Qaqish, B.F. (1996). Deletion diagnostics for generalized estimating equations. *Biometrika*. 83, 551-562.
- [92] Prentice, R.L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* 44, 1033-1048.

- 
- [93] Price, W.E., Grizzle, J.E., Postlethwait, R.W., Johnson, W.D. y Grabicki, P. (1970). Result of operation for duodenal ulcer. Sugery, *Gynecology and Obstetrics* 131, 233-244.
- [94] Purdie, D., Dobson, A. y Williams, G. (2003). *Applications of generalized estimating equations to longitudinal data*. Longitudinal Studies, The University of Queensland.
- [95] Rao, J.N.K. y Scott, J. (1981). The Analysis of Categorical Data From Complex Sample Surveys: Chi-Squared Test for Goodness of Fit and Independence in Two-Way Tables. *Journal of the American Statistical Association*, 76, (374), 221-230.
- [96] Rao, J.N.K. y Scott, J. (1984). On chi-squared Test for Multi-way Contingency Tables with Cell Proportions Estimated from Survey Data. *Ann. Statist.*, 12, 46-60.
- [97] Rao, J.N.K. y Scott, J. (1987). On Simple Adjustments to Chi-Square Tests with Sample Survey Data. *The Annals of Statistics*, 15, (1), 385-397.
- [98] Roberts, G., Rao, J.N.K. y Kumar, S. (1987). Logistic Regression Analysis of Sample Survey Data. *Biometrika*, 74, (1), 1-12.
- [99] Rotnitzky, A. y Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear for cluster correlated data, *Biometrika*, 42, 311-323.
- [100] Ruppert, D. y Carroll, R.J. (2000). Spatially-adaptive penalties for spline fitting. *Australian & New Zealand J. Statist.*, 42, (2), 205-223.
- [101] Self, S.G. y Liang, K.Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Test Under Nonstandard Conditions. *Journal of the American Statistical Association*, 82, (398), 605-610.
- [102] Shin, D.W., Park, C.G. y Park, T. (1996). Testing for ordered group effects with repeated measurements. *Biometrika*, 83, (3), 688-694.

- 
- [103] Skellam, J.G. (1948). A probability distribution derived from the binomial distribution regarding the probability of success as a variable between sets of trials. *J. Roy. Stat. Soc. Ser. B.* 10, 257-261.
- [104] Stiratelli, R., Laird, N. y Ware, J.H. (1984). Random-Effects Models for Serial Observations with Binary Response. *Biometrics*, 40, (4), 961-971.
- [105] Stokes, M. E., Davis, C. S. y Koch, G. C. (2000). *Categorical Data analysis Using The SAS<sup>®</sup> System, 2nd Edition*. Cary, NC: SAS Institute Inc.
- [106] Stram, D.O. y Lee, W.L. (1994). Components Testing in the Longitudinal Mixed effects Model. *Biometrics*, 50, (4), 1171-1177.
- [107] Tan, M., Qu, Y. y Kutner, M.H. (1997). Model diagnostics for marginal regression analysis of correlated binary data. *Communication in Statistics Simulation*, 26, 539-558.
- [108] Thall, P.F. y Vail, C.S. (1990). Some covariance models for longitudinal count data. *Biom. J.* 46, 441-452.
- [109] Toon, T. (2000). *A Primer in Longitudinal Data Analysis*. Sage, Thousand Oaks, CA.
- [110] Venezuela, M. K., Botter, D. A. and Sandoval, M.C. (2007). Diagnostic techniques in generalized estimating equations. *Journal of Statistical Communication and Simulation*, 77, (10), 879-888.
- [111] Verdeke, G y Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- [112] Wedderburn, R.W.M. (1974). Quasi-likelihood funtions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61, (3), 439-447.
- [113] West, B.T., Welch, K.B. y Galecki, A.T. (2007). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman & Hall.
- [114] Williams, D.A. (1987) Generalized Linear Model Diagnostics Using the Deviance and Single Case Deletions, *Applied Statistics*, 36, 181-191.

- 
- [115] Williams, O.D. y Grizzle, J.E. (1972). Analysis of contingency tables having ordered response categories. *Journal of the American Statistical Association* 67, 55-63.
- [116] Zeger, S. L. y Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42, 121-130.
- [117] Zeger, S. L., Liang, K. Y. y Albert, P.S. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, 44, 1049-1060.
- [118] Zhao, L.P., Prentice, R.L. (1989). Correlated binary regression using a quadratic exponential model. *Biometrika* 77, 642-648.
- [119] Zorn, C.J.W. (2001). Generalized Estimating Equation Models for Correlated Data: A Review with Applications. *American Journal of Political Science*, 45, (2), 470-490.

